

とわかった。これは、クロス表の分析と矛盾しない。

表 11 趣味・教養活動の参加の規定要因

	B	標準誤差	Wald	自由度	有意確率	Exp (B)	
男性ダミー	-.444	.036	153.377	1	.000	.642	***
年齢	.035	.006	30.231	1	.000	1.035	***
中卒ダミー	-.750	.045	272.394	1	.000	.472	***
専門学校卒ダミー	.151	.063	5.812	1	.016	1.163	*
短大卒ダミー	.560	.075	55.548	1	.000	1.751	***
大卒ダミー	.663	.050	177.782	1	.000	1.941	***
大学院卒ダミー	.857	.191	20.182	1	.000	2.357	***
学歴その他ダミー	.055	.217	.064	1	.800	1.056	
配偶者ありダミー	.211	.048	19.676	1	.000	1.235	***
介護ありダミー	-.197	.050	15.556	1	.000	.821	***
困難ありダミー	-.339	.056	36.233	1	.000	.713	**
仕事ありダミー	-.163	.048	11.828	1	.001	.849	**
収入	.000	.000	3.712	1	.054	1.000	+
持家ありダミー	.648	.056	135.179	1	.000	1.912	***
社宅ダミー	.778	.198	15.441	1	.000	2.176	***
住宅その他ダミー	.083	.135	.377	1	.539	1.087	
定数	-1.696	.390	18.920	1	.000	.183	***

*** p<0.001 ** p<0.01 * p<0.05 +<0.1
 変数のそろった18000ケースを分析
 R2=0.051

表 12 スポーツ・健康活動の参加の規定要因

	B	標準誤差	Wald	自由度	有意確率	Exp (B)	
男性ダミー	.085	.033	6.487	1	.011	1.089	*
年齢	.040	.006	46.911	1	.000	1.041	***
中卒ダミー	-.601	.046	171.513	1	.000	.548	***
専門学校卒ダミー	.050	.057	.752	1	.386	1.051	
短大卒ダミー	.281	.062	20.318	1	.000	1.325	***
大卒ダミー	.538	.044	148.044	1	.000	1.712	***
大学院卒ダミー	.732	.164	20.047	1	.000	2.080	***
学歴その他ダミー	-.080	.206	.150	1	.699	.923	
配偶者ありダミー	.171	.045	14.106	1	.000	1.186	***
介護ありダミー	-.097	.047	4.301	1	.038	.907	*
困難ありダミー	-.660	.056	139.458	1	.000	.517	***
仕事ありダミー	-.457	.044	108.240	1	.000	.633	***
収入	.000	.000	4.781	1	.029	1.000	*
持家ありダミー	.391	.056	48.826	1	.000	1.479	***
社宅ダミー	.337	.180	3.498	1	.061	1.400	+
住宅その他ダミー	.088	.136	.416	1	.519	1.092	
定数	-2.448	.365	45.086	1	.000	.086	***

*** p<0.001 ** p<0.01 * p<0.05 +<0.1
 変数のそろった18002ケースを分析
 R2=0.038

スポーツ・健康活動

男性、年齢が高い、学歴が高い（専門学校卒、短大卒、大卒、大学院卒の順に高卒に比べて参加ありが増加する）、配偶者がいる、収入が高い、持家や社宅であるといった点が、参加に正の効果をもたらす。逆に、中卒、介護あり、生活に困難あり、仕事ありだと、参加に負の効果があるとわかった。これもまた、クロス表の分析と矛盾しない。

6. 6. まとめ

趣味・教養やスポーツ・健康にかかわる活動に参加することは、よりよい精神状況、飲酒習慣、喫煙習慣、運動習慣、健康状態と関係していることがわかった。因果関係は厳密には特定できないが、「自分のため」の活動に取り組んでいる人は、よりよい生活を送っている。

そのような活動への参加を促す要因は、「趣味・教養」は女性、「スポーツ・健康」は男性、前者は治療費用なし、後者は治療費用ありという違いがある以外は、大きな傾向は同じである。すなわち、年齢が高く、学歴が高いこと。配偶者がいるが、それ以外の同居者は少ないこと、介護の必要がないこと。自身の健康状態がよく、日常生活で活動の困難がないこと。仕事をしていないこと。仕事をしている場合は、日々の生活のために働いているわけではなく、収入的に安定し、日程的に余裕がある働き方をしていること。収入源が安定的な不労所得であること。年齢がいったら、仕事を辞めたいと思っていること。住宅が持家や社宅であること、以上のような傾向が見られる。

このような結果は、常識的な予測とほぼ一致しているといえよう。これは、裏を返せば、このような条件が得られない人は、自分のための余暇の活動に参加できないということを意味している。あたりまえのことであるが、さまざまな生活上の正の効果をもたらす余暇活動を、多くの中高年者が享受できるような、雇用環境や社会保障の整備が必要であろう。そのことが、さらに、より他人や社会のためのボランティアな活動をも活発化させる第一歩と考えられる。

Ⅲ. 資 料

2012年8月8日(水)
社人研第4会議室
13:30-15:30
厚労科研パネルプロジェクト④

伝達研修

傾向スコア分析

(Propensity Score Analysis)

鎌田健司
(社人研人口動向研究部)

伝達研修

- 研修に参加してきた内容を報告する。
 - 「傾向スコア(propensity score)その考え方と実際」(講師:岩崎学-成蹊大学-),「因果のメカニズムを解きほぐす」,統計数理研究所,2010年9月13~14日.
 - 「Propensity Score分析」(講師:高橋理-聖路加国際病院-),SPSS統計セミナー「臨床研究におけるPropensity score(傾向スコア)推定法とそれを用いた解析のためのSPSS活用法」,IBM箱崎本社,2012年6月28日.

主要テキスト

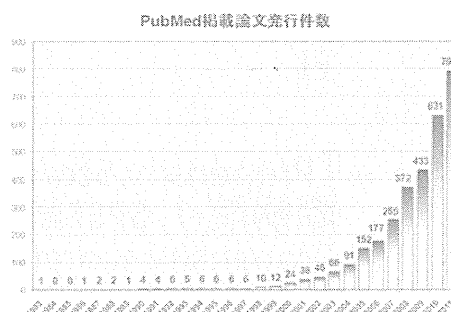
- 星野崇宏(2009)『調査観察データの統計科学 因果推論・選択バイアス・データ融合』(確率と情報の科学シリーズ), 岩波書店.
 - Guo, S. and Fraser, M. W., (2010) *PROPENSITY SCORE ANALYSIS*, SAGE Publications, Inc.
- 両書籍には、RやSTATAを用いた解析例が提示されているため、実際に推定する際に参考になる。

②

傾向スコアとは

- 傾向スコア分析は、統計的因果推論の議論の中で考案された手法
- 1983年に考案される。(Rosenbaum and Rubin 1983)
- 医学分野で発展。
- 大規模調査とネット調査を結合する手法等にも応用(星野 2003,2009)

* 高橋資料(スライド24)より→



Rosenbaum, P. R. and Rubin, D. B., 1983. "The Central Role of the Propensity Score in Observational Studies for Causal Effects", *Biometrica*, Vol. 70, pp. 41-55.

星野 崇宏(2003)「調査データに対する傾向スコアの適用」, 『品質』Vol.33, No.3, pp.44-51.

③

縦断調査を用いた分析に どのように使えるのか？

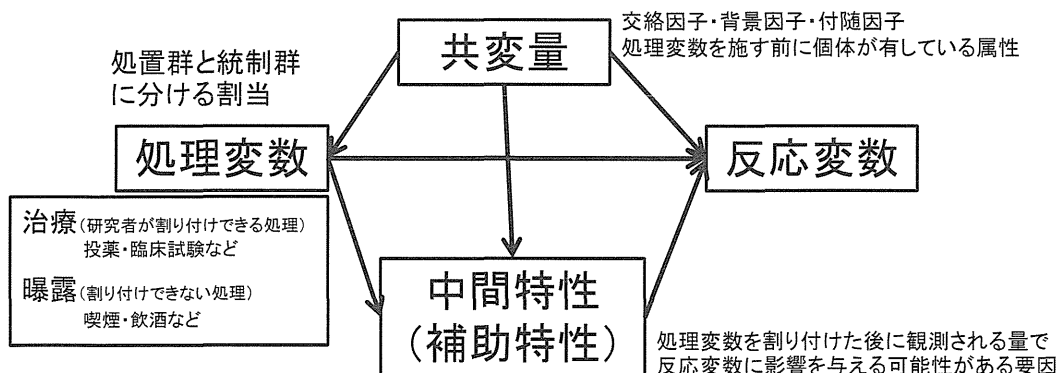
- 政策効果(制度の利用の有無等)の測定
 - 「処理」変数の選択
 - 政策効果の因果効果を推定・検定
- 標本の脱落が目的変数に及ぼす因果効果
 - 脱落を処理とした場合の目的変数への因果効果
 - 「ランダムな脱落」(脱落するかどうか脱落時あるいは脱落後に得られる観測値に依存しない場合)
 - (1)脱落する確率に影響を与える変数と従属変数との回帰関係が明確な場合: 回帰分析モデルを最尤推定
 - (2)脱落するかどうかのモデリングが明確な場合: 傾向スコアを利用してIPW推定量や二重にロバストな推定量を計算する
 - 「ランダムでない脱落」
 - Diggle and Kenward(1994)モデル
 - ヘックマンのプロビット選択モデル

(星野 2009)

4

統計的因果推論

- 統計的因果推論の主目的は、処理変数が反応変数に及ぼす因果関係を定量的に評価し、それを利用すること

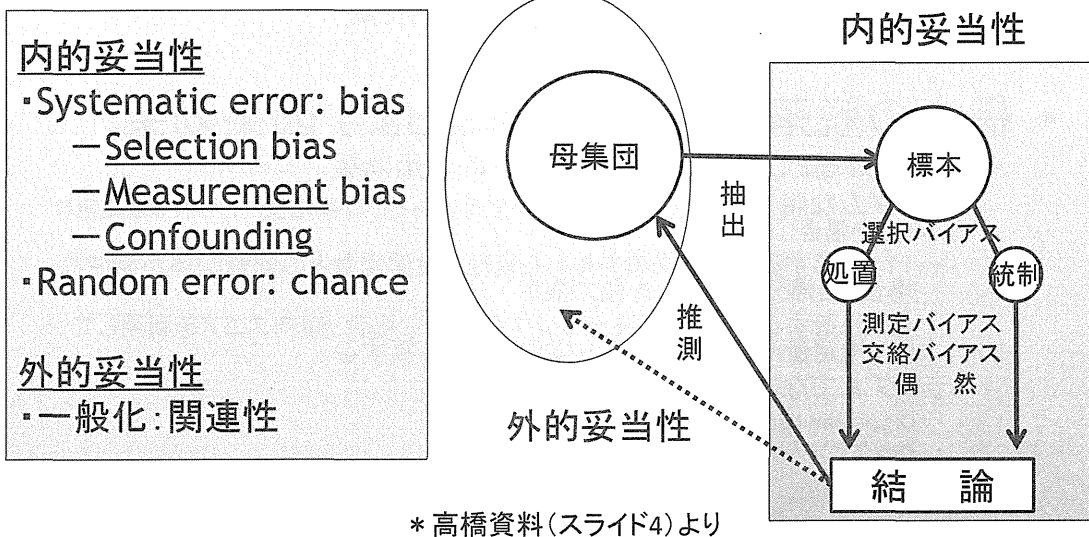


宮川 雅巳(2004)『統計的因果推論—回帰分析の新しい枠組み—』朝倉書店

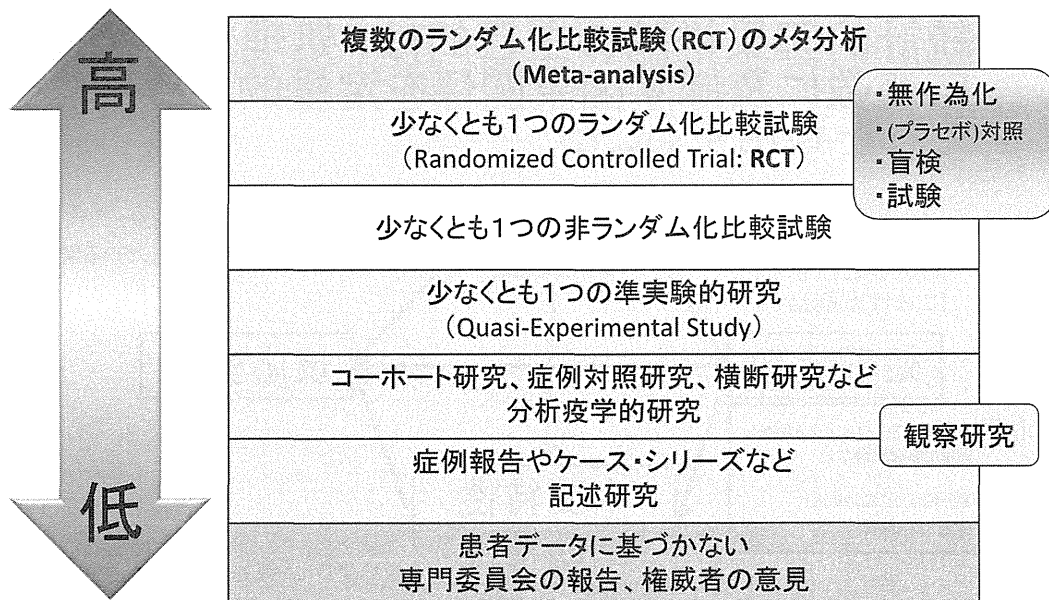
5

バイアスを如何に処理するか？

- 真の因果関係の測定のためには、バイアスを如何に処理するかが問題となる。



エビデンス・レベル (臨床研究の例)



ランダム化比較研究

- 処置群・統制群の割り当てがランダム
 - 全ての既知および未知の予後因子に関して処置群と統制群が同等であることを保証する。

- 利点

- 治療効果や曝露を調整できる。
- バイアスを極力除去できる。
- 科学的に妥当な結果を得られる。

実験的研究
Experimental
Study

- 欠点

- 倫理的問題(統制群の患者に本当に処置しなくていいのか?)
- 時間と費用(科学的な実験計画にはコストがかかる)
- 処置(治療)が日常的に行われている場合。
- 外的妥当性が乏しい(共変量の調整が断片的)。

8

観察研究: コーホート研究

- 処理などの介入をせず、観察した結果をもとに分析する。横断調査による回顧(retrospective)や縦断調査による事象の発生(prospective)を捉える。

- 利点

- 倫理的な問題で実行困難なことが少ない。
- 外的妥当性が得られやすい(多くの共変量を調整)。

- 欠点

- 治療効果の測定においては、治療が非ランダムに割り付けられている。
- 様々な交絡因子(Confounding)が存在する。
- 観察できない交絡因子の調整ができない。

9

両研究デザインの良い所取り

- それが、傾向スコア分析。
- 簡単に言うと…
 - 観察研究に用いられる標本調査等を用いて、
 - 実験的研究を行う分析手法。

共変量の
バランス

傾向スコア分析の長所

- 多くの共変量を調整できる。
 - 特に、イベント数(目的変数)に比べて共変量が多いとき(1:7)。
- 処理群と統制群のバランスを保つため、共変量の問題に直接的に対応。
- 処理効果の推定は傾向スコアモデルにおける定式化の誤りによる影響をあまり受けない。

傾向スコア分析は2段階の手順

傾向スコア
の推定

分析

高橋資料(スライド23,25)より

10

傾向スコアの定義

- 複数の共変量を用いて2群(処置群・統制群)に割り当てられる確率を予測するスコア(高橋資料)
- 測定された共変量のベクトルによって得られた特定の処置(treatment)に割り当てられる条件付き確率(Guo and Fraser, p.132)
- 従属変数と独立変数間の因果関係を推定するために、それ以外の変数で従属変数に関連する要因=共変量を一つの変数に縮約したスコア

$$e(x_i) = \text{pr}(W_i = 1 \mid X_i = x_i)$$

i: 標本(i=1, ..., N), $e(x_i)$: 傾向スコア, $W_i = 1$: treatment(処置),
 $W_i = 0$: nontreatment(非処置=統制), X_i : 共変量ベクトル, x_i : 観察される共変量
ただし, $0 \leq e(x_i) \leq 1$.

このスコアを用いて処置群と統制群をマッチングさせることで、処置群と統制群は共変量が同じになることから、独立変数の従属変数に対する因果効果を推定することができる。

11

具体的な工程(Guo & Fraser 2010)

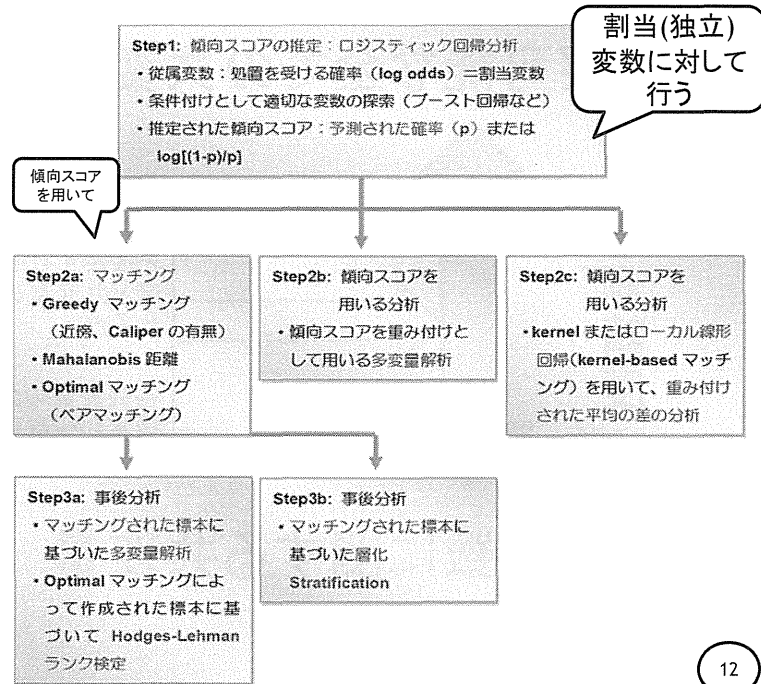
Step 1 処置群と統制群の間の不均等 (imbalance) を生じさせると推測される最良な条件変数や従属変数に関連する共変量を探す。

→ロジスティック回帰モデル(または、プロビット回帰、判別分析等)または多項ロジットモデルを用いて、処置の効果を分析

Step 2a マッチングまたは再サンプリング (re-sampling)
オリジナルサンプルは処置群と統制群の間の共変量は均等ではないが、傾向スコアに基づいた再サンプリングは測定された処置のサンプリングバイアスを統制し、測定された共変量を均等にします。

Step 3a マッチングされた標本を元にしたマッチング後の分析

Step 3b 傾向スコアの層化 (stratification) を用いて事後マッチング分析



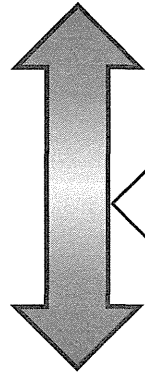
Step 1 傾向スコアの推定

- 傾向スコアは割当変数(治療、喫煙、脱落等)を従属変数として、真に知りたい従属変数(死亡、出生等)に関連のある共変量を独立変数として二項ロジスティック回帰(プロビットモデル)によって推定する。
 - 割当変数が3値以上であれば、多項ロジット・順序ロジット等を用いる。
- 傾向スコアが同じ対象は、共変量(交絡因子)の分布が同じになる傾向がある→共変量が均衡(バランス)⇔RCT
- 傾向スコアの推定に用いる共変量選択には諸説(議論)がある。
 - 観察される全ての共変量を投入する(岩崎)。
 - 真に知りたい従属変数に関する共変量を投入する(高橋)。
 - 臨床的意義・欠損値が少ない・単変量解析で関連する
 - イベント数は統制する共変量の7~10倍必要(臨床研究)→ステップ・ワイズ法で調整

共変量の選択についての注意点

- 観察可能な全ての共変量

高
予測精度
低



多くの変数を投入して傾向スコアを推定する場合、多ければ多いほど共変量のバランスは担保されるが、

各共変量の欠損値などによって標本が少なくなる可能性もある。

- 目的変数に関する共変量

14

良い傾向スコアかどうかをどのように判断するのか？

- 前提条件

- 強く無視できる割り当て
- 測定されていない共変量はない

「どちらの群に割られるかは観測された共変量の値に依存し、従属変数の値の高低によっては依存しない」
(星野・繁樹 2004)

- 前提条件のチェック

- モデルの予測度を確認する
 - C値(AUC) > 0.8
- 共変量が群間で差がないかを示す
 - 2群間の検定

15

* 高橋資料(スライド44-45)

星野 崇宏・繁樹 算男(2004)「傾向スコア解析法による因果効果の推定と調査データの調整について」、『行動計量学』第31巻第1号, pp.43-61.

良い傾向スコアかどうかをどのように判断するのか？

- Hosmer-Lemeshow検定
 - ロジスティック回帰モデルへの適合度を調べる統計学的検定。
 - 観測された事象率がモデル母集団のサブグループでの期待される事象率に適合するかどうかを評価。
- ROC(Receiver Operating Characteristic)曲線と
 - 観測された信号からあるものの存在を判定する際の基準となる特性。
 - 有効であれば、この曲線は45度の線から左上に離れる。離れば離れる程、検査として有効。
- AUC(Area under the curve)値
 - 割当ての当てはまりの良さを示す。 0.7~0.8が良い
 - 0から1までの値をとり、完全な分類が可能となるとき面積は1で、ランダムな分類の場合は0.5

16

傾向スコアモデルのチェック

- 唯一の問題は、共変量のバランスの確保
- 傾向スコアを推定する際には、オーバーフィッティング・共線性・パラメータの標準誤差が大きい等は問題ではない。
- バランスさえ確保されていれば、それでよい。
- 例) 脱落を割当て変数としたときの傾向スコアの共変量のチェック * 高橋資料(スライド41)より
 マッチング後に脱落の有無での共変量の差がなくなっている(t検定)

脱落要因	脱落前				脱落後			
	平均 (脱落)	平均 (非脱落)	平均比 (脱落/非脱落)	t検定	平均 (脱落)	平均 (非脱落)	平均比 (脱落/非脱落)	t検定
学歴：中学校	0.02479	0.01679	1.46660		0.02479	0.02706	0.91651	
学歴：高校卒	0.28889	0.27280	1.03220		0.28889	0.30270	0.95528	
学歴：専門学校卒	0.19126	0.18557	1.03450		0.19126	0.18170	1.04330	
学歴：短大・高専卒	0.21608	0.25401	0.84475	*	0.21608	0.22001	0.98700	
学歴：大学・大学院卒	0.28099	0.27079	1.02420		0.28099	0.26853	1.02880	
職業：中小企業(従業員500人以下)	0.27961	0.32022	0.82658	*	0.27961	0.28227	0.99458	
職業：大企業(従業員500人以上)	0.14876	0.13148	1.11000		0.14876	0.14849	1.00150	
職業：自営/フリー	0.02361	0.02704	0.87713		0.02361	0.02486	0.95089	
職業：自営業・会社役員	0.03188	0.02965	1.07360		0.03188	0.03756	0.85366	
職業：非正規雇用	0.34238	0.34670	0.99507		0.34238	0.33970	1.00380	
職業：学生	0.02361	0.03059	0.77835		0.02361	0.02463	0.95206	
職業：無業	0.08028	0.08173	1.27610	+	0.08028	0.06977	1.13770	
職業：その他	0.06906	0.05259	1.30190	+	0.06906	0.07250	0.96376	
収入階層(対数化)	2.94970	3.05980	1.17220	**	2.94970	2.98100	1.00589	
収入の階層(対数化)	5.04790	4.85640	1.25650	**	5.04790	5.02290	1.10580	
年齢と性別	0.57261	0.76762	1.37330	**	0.57261	0.56925	0.98936	
年齢と性別	0.12515	0.11805	1.05260		0.12515	0.13146	0.95890	
年齢と性別	0.30224	0.11414	2.68790	**	0.30224	0.29929	1.00580	
結婚意欲：絶対しない	0.34711	0.34276	1.00700		0.34711	0.34331	1.00520	
結婚意欲：どちらかといえほしい	0.33884	0.36069	0.97250		0.33884	0.35312	0.98075	
結婚意欲：どちらともいえない	0.23731	0.21410	1.07670		0.23731	0.22945	1.02370	
結婚意欲：あまりたくない	0.05549	0.06472	0.86877		0.05549	0.05143	1.07430	
結婚意欲：絶対しない	0.02125	0.01772	1.19640		0.02125	0.02289	0.93812	

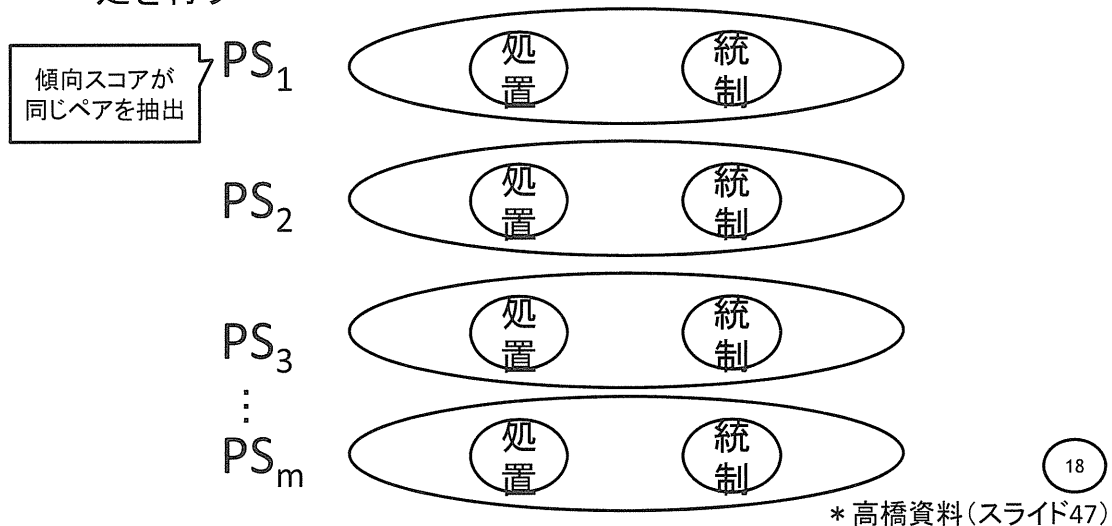
(鎌田 2011報告書より)
 統計ソフトR
 Matchingパッケージ
 MatchBalance()により
 検定

17

鎌田 健司(2011)「第8章 傾向スコア・脱落サンプルバイアスの検定法」, 厚生労働科学研究費補助金『パネル調査(縦断調査)に関する統合的高度統計分析システムの開発研究』(研究代表者:金子隆一), 平成22年度総括研究報告書。

Step2a マッチング (Matching)

- 傾向スコアを用いて処置と統制でペアを抽出
(マッチングされないデータは除外されてしまう)
- 処置群と統制群のペア間の差の平均を用いて因果効果の推定を行う



Step3a 多変量解析

共変量の影響が調整されたデータセット

- マッチングを行ったデータセットを用いた多変量解析を行う
 - 対応のあるデータにおける条件付き推定
- 手法: 傾向スコアを多変量解析のモデルに投入
 - 回帰分析、ロジスティック回帰分析など
- 目的変数: アウトカム・イベント
- 独立変数: 処理変数 + 傾向スコア (+ 重要な共変量)

Step3a 多変量解析
分析事例:

院外での心停止患者の入院前におけるエピネフリンの使用と生存

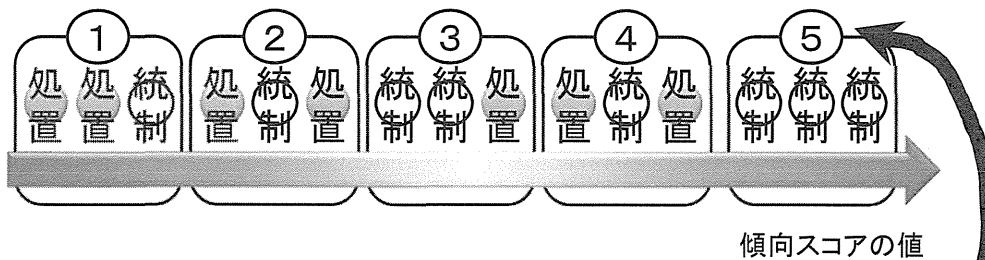
- 傾向スコア分析を用いた理由: RCTモデルが使用できないため
- データ: 2005年1月1日から2008年12月31日までの心停止患者の症例 (エピネフリン使用15,030人、不使用402,158人)
- 分析手法: 傾向スコアを用いた条件付きロジスティック回帰分析
- 共変量: 発生年、年齢、性別、発症時目撃者有無、発症時家族目撃有無、心停止の主因、心肺機能蘇生処置有無、救急による処置(救急車内での処置、救急を呼んでから到着までの時間、病院までの時間、電気ショック使用有無など)など
- 結果:
 - 対応のないデータ : オッズ比3.75 (3.59-3.91), N=417,155
 - 対応のないデータ(+選択された共変量) : オッズ比3.06 (2.93-3.21), N=412,078
 - 対応のないデータ(+全ての共変量) : オッズ比2.36 (2.22-2.50), N=391,046
 - 対応のあるデータ : オッズ比1.91 (1.78-2.05), N= 26,802
 - ⇒ - 対応のあるデータ (+傾向スコア+選択された共変量) : オッズ比2.24 (2.03-2.48), N= 26,802
 - ⇒ - 対応のあるデータ (+傾向スコア+全ての共変量) : オッズ比2.51 (2.24-2.80), N= 26,802

20

Hagihara et al. (2012) "Prehospital epinephrine use and survival among patients with out-of-hospital cardiac arrest", The Journal of the American Medical Association, 307(11), pp.1161-1168.

Step3b 層化 (Stratification)

- 標本を傾向スコア順に並び換え、標本サイズが同じ約5つの群に分ける



各群内で処置群と統制群を比較し、それを基に処理効果を評価する (統制群と処置群の結果の平均の差を比較、対応のあるMantel-HentelのOdds比を計算するなど)

利点: 傾向スコアをもつ全ての標本が対象となる

欠点: 標本数が少ない場合、処置が含まれない群が生じる可能性

21

* 高橋資料(スライド49)

Step 2b,2c 重み付け (Weighting)

- 傾向スコアを多変量解析モデルの重み付けとして用いる(マッチングはしない)
- IPW(inverse-probability score-based weighted methods) 推定量を用いた重み付け (星野 2009、坂本 2006)

22

坂本 和靖(2006)「サンプル脱落に関する分析—「消費生活に関するパネル調査」を用いた脱落の規定要因と推計バイアスの検証」、『日本労働研究雑誌』No.551, pp.55-70.

傾向スコア分析の問題点

- 測定されていない共変量は調整できない
- 共変量のバランスをチェックするための方法論は確立されていない(一律の方法がない)
 - 共変量の選択法
- ベストな傾向スコアモデルを決める明確な基準がない
 - 傾向スコアモデルの予測精度とマッチング時の傾向スコアの使い方によって、その後の分析結果に変動が生じる
 - 分析サンプル数の確保と予測精度の関係

23

おわりに

- 傾向スコア分析は観察研究において、処理効果(治療、政策)を比較したい場合に利用
 - 特に共変量が多く、イベントが少ないとき
(おおよそ、共変量1:イベント7ぐらいの標本)
- Stata,R等の統計ソフトを用いることで、比較的容易に分析が可能
- 調査データでも統計的な因果効果の測定が可能となるため、さまざまな分野で利用されていく可能性が高い

24

(参考)傾向スコアの特徴(Guo & Fraser 2010)

- 傾向スコアの利点は、マッチングにしても層化にしても、次元を減少させることができることにある。傾向スコア x_i には多くの共変量が含まれているが、1つの次元で表現できる。ただし共変量が多くなるにつれて、マッチングさせる対象が少なくなる。
 - 1) 傾向スコアが同じであるならば、処置群と統制群の共変量は同じ。
 - 2) 傾向スコアが同じであるならば、処置群の割当と共変量は独立。
→傾向スコアが同じであることは、無作為に割当られた実験下において処置が割当られる確率が同じ。
 - 3) もし強く無視された割当の仮定が保持され $e(x_i)$ がバランシングスコアであるならば、2つの割当(処置群と統制群)の期待差(expected difference)はATE (Average Treatment Effect)と同じ。
$$E[E(y_1, |e(x_i), z=1) - E(y_0, |e(x_i), z=0)] = E[y_1 - y_0 | e(x_i)]$$
 - 4) 傾向スコアを用いた3つのアプローチがある。
 - a) ペアマッチング(Pair matching): 傾向スコアが同じ条件における、処置群と統制群の応答の期待される差は、ATEと同じ。
 - b) 層化(Stratification, Subclassification): 傾向スコアの値をいくつかの層(subclass)に分けて、ATEを計算する。
 - c) 共変量調整(Covariance adjustment): 傾向スコア $e(x)$ において処置の割当が強く無視できると仮定できるとき、 $Y(t=0,1)$ の条件付き期待は線形関係となる。

(参考) 傾向スコアを用いると 因果効果が推定できる理由 (星野 2009)

- 傾向スコアを用いた共変量調整は、ランダムな割当がされているとすれば、全ての共変量を用いて調整を行ったと同じだけの偏りを減少させることができる。
- 因果効果: $E(y_1) - E(y_0) = E(y_1 - y_0) = E_e(E(y_1 - y_0 | e))$
- さらに、 $(y_1, y_0) \perp z | b(x)$ より、
- 結果として因果効果は $E(y_1) - E(y_0) = E_e[E(y_1 | e, z = 1) - E(y_0 | e, z = 0)]$
- 傾向スコア e を所与として (同じ傾向スコアの値の元で)、各群から得られた y_1, y_0 から $y_1 - y_0$ を計算し、それを傾向スコアの分布で期待値を取ると、因果効果になる。
- 右辺で、 y_1 については $z=1$ 、 y_0 については $z=0$ が所与となっており、潜在的な結果変数のうち観察されていない値については考えなくてもよいということになるため、傾向スコアと観察されている従属変数の情報を用いれば因果効果の推定が可能。

y_1 : 従属変数(応答 = 1)
 y_0 : 従属変数(応答 = 0)
 e : 傾向スコア
 z : 割当
 $b(x)$: バランシングスコア
 x : 共変量

(参考) 統計ソフト・関数一覧 (Guo & Fraser 2010)

手法	Stata	R
Heckman (1978, 1979) sample selection model	heckman (StataCorp 2003)	sampleSelection (Toomet&Henningsen 2008)
Maddala (1983) treatment effect model	treatreg (StataCorp 2003)	
Rosenbaum&Rubin(1983) propensity score matching	psmatch2 (Leuven&Sianesi 2003)	cem (Dehejia&Wahba 1999, Iacus&King&Prro 2008) Matching (Sekhon 2007) MatchIt (Ho&Imai&King&Stuart 2004) PSAgraphics (Helmerich&Pruzek 2008) Whatif (King&Zeng 2006, 2007) USPS (Obenchain 2007)
Generalized boosted regression	boost (Schonlau 2007)	gbm (McCaffrey&Ridgeway&Morral 2004)
Optimal matching (Rosenbaum 2002)		optmatch (Hansen 2007)
Postmatching covariance imbalance check (Haviland& Nagin&Rosenbaum 2007)	imbalance (Guo 2008)	
Hodges-Lehmann aligned- rank test after optimal matching (Haviland&Nagin &Rosenbaum 2007, Lehmann 2006)	hedgesl (Guo 2008)	
Matching estimators (Abadie&Imbens 2002, 2006)	nnmatch (Abadie&Drukker&Herr&Imbens 2004)	Matching (Sekhon 2007)
Kernel-based matching (Heckman&Ichimura&Todd 1997, 1998)	psmatch2 (Leuven&Sianesi 2003)	
Rosenbaum (2002) sensitivity analysis	rbounds (Gangl 2007)	rbounds (Keele 2008)

