

ロジスティック回帰分析と傾向スコア (propensity score) 解析

大林 準*

天理よろづ相談所 医学研究所

受付 2016/4/13, 受理 2016/7/27, online 公開 2016/12/25

医学統計において、多変量解析で用いられる統計手法の一つにロジスティック回帰分析がある。ロジスティック回帰分析は、目的変数が、「生存・死亡」や「陽性・陰性」といった名義変数の場合に用い、治療（介入）の効果について、目的変数に関わる因子（共変量）が回帰式にどの程度関与しているかを解析するものである。ロジスティック回帰モデルでは、その結果に対する確率を P とし、共変量を $x_1, x_2, \dots$ とした場合、 $\log(P/(1-P)) = b_0 + b_1x_1 + b_2x_2, \dots$ といった式で表せる。無作為化されていない後ろ向き研究においては、交絡因子とされるものが、解析結果に対して影響を及ぼすことがたびたび見られる。交絡因子の調整方法として、近年、傾向スコア (propensity score) 解析が提唱され、この傾向スコアには、ロジスティック回帰分析で求めた予測確率を用いる。傾向スコア解析とは、潜在的な交絡要因となる様々な共変量を、傾向スコアという一つの合成変数に縮約（一次元化）し、その傾向スコアを基準として、交絡因子の影響を除去するためにマッチングや層別化等を行うものである。

キーワード: ロジスティック回帰分析, 傾向スコア (propensity score), 交絡因子

1. はじめに

臨床研究を行う際、目的変数と目的変数に関わる因子についての関係を調査・解析する方法として、単変量解析 (univariate analysis) と多変量解析 (multivariate analysis) がある。

単変量解析とは、目的変数に関わる因子（以後、共変量とする）が 1 つの場合の解析である。平均値や標準偏差などの基本統計量、 t 検定、カイ 2 乗検定、ログランク検定（カプランマイヤー曲線）等が単変量解析になる。共変量または目的変数が 2 つ以上ある場合は、多変量解析を用いる。例えば、目的変数が「ある重篤な疾患で緊急手術をした患者の ICU 退室時の生存・死亡」とした時に、「年齢・性別・BMI・発症から手術までの時間・手術

中の輸血量」の 5 つの共変量が、ICU 退室時の生存・死亡にどれ程影響しているかを“同時に”解析したい場合に用いる。

多変量解析に用いるいくつかの共変量は、目的変数に直接関わるだけでなく、共変量間でも影響し合っている。これを交絡といい、交絡した因子は、交絡因子とされ、しばしば、解析の妨げになる。¹ この交絡因子の影響を効果的に取り除くための手法として、1983 年に Rosenbaum と Rubin によって傾向スコア (propensity score) 解析法が初めて提案され、² 近年使用されることが多くなった。³

ロジスティック回帰分析の結果の解釈や、傾向スコア解析の概念・使用方法について紹介する。

2. ロジスティック回帰分析

1) 多変量解析

多変量解析は、共変量または目的変数が 2 つ以上ある場合に用いる解析法である。多変量解析に

* 連絡先: 大林 準
〒632-8552 天理市三島町 200
天理よろづ相談所 医学研究所
e-mail: oobayasi@tenriyorozu.jp

は様々な方法があるが、医学統計では、重回帰分析・ロジスティック回帰分析・Cox 比例ハザードモデルの3つの方法がよく使用される。解析方法の選択は、目的変数の種類や時間的要素のありなしにより決定される(表1)。⁴⁶⁾ 本稿では、目的変数が「あり・なし」のように2値である場合のロジスティック回帰分析について解説する。

ロジスティック回帰分析において、「あり・なし」の目的変数のうち、“あり”をイベントとした場合の、イベント(あり)の出現する確率を P とした場合、次の式で表せる。

ロジスティック回帰モデル

$$P = \frac{1}{1 + \exp\{-(b_0 + b_1 x_1 + b_2 x_2 + \dots + b_p x_p)\}}$$

$$\ln\left(\frac{P}{1-P}\right) = b_0 + b_1 x_1 + b_2 x_2 + \dots + b_p x_p$$

P : イベントの出現する確率, b_0 : 定数, b_p : 偏回帰係数, x_p : 共変量

2) 共変量と交絡因子

ロジスティック回帰モデルにおいて、「治療効果あり・治療効果なし」や「陽性・陰性」といった目的変数に対して、関連する因子は共変量(説明変数・独立変数)と言う。共変量は連続変数と名義変数が混在しても構わない。また、共変量は、交絡因子である場合とない場合がある。

交絡因子とは、「調べようとする因子以外の因子で、結果(目的変数)に影響を与えるもの」であり、交絡因子であるための3つの条件として、①交絡因子は問題となっている目的変数(疾患の発生のあり・なし、死亡の有無等)に影響を与える可能性のある因子である。②問題となっている要因(原因)と関係がある。③問題となっている要因と結

果の因果連鎖の中間変数ではない。という条件が必要である。

例えば、肺がんの発生の有無(目的変数)と飲酒についての関係について調べたいとする。その際に、「飲酒あり」の群に肺がんの発生が多かったとする。これは、飲酒により肺がんが引き起こされた結論づけでよいだろうか。この場合、飲酒以外の他の因子が肺がんの発生に関与していることのチェックも必要である。

「喫煙」という因子があった場合、喫煙は肺がん発生に影響を与える可能性のある因子であり(条件①)、飲酒する人はしない人よりも喫煙する率が高い(条件②)が、飲酒の結果が喫煙ではない(条件③)、このような場合、喫煙は、飲酒の有無と肺がんの発生の関係を調べるときの交絡因子と言える(図1)。

さらに、目的変数に単に影響するだけでなく、悪い意味で解析の結果を歪めるおそれのある因子も交絡因子である。例えば冠不全が「男性」に多いので男であることが冠不全の一因と解釈したとする。ところが本当は「喫煙」が原因であり、男性に喫煙者が多い(性と関係がある)ので男性に冠不全が多いような歪んだ結果を生むこともあり得る。このような場合は、喫煙が交絡因子で、解析結果を歪めることになる。どちらが正しいかを調べるためには、喫煙者と非喫煙者をひっくりめて丸ごと比較するのではなく、別々の層にわけ、それぞれの層の中で冠不全の発生に男女差があるかをみればよい。つまり喫煙に関して似たもの同士を比較する。これが層別解析である。後述する傾向スコア解析の基礎となっている。

表 1. 医学統計でよく用いられる多変量解析

アウトカムの種類(例)	時間的要素	解析方法
連続変数 (出血量等の数値)	なし	重回帰分析
名義変数 (改善あり・改善なし)	なし	ロジスティック回帰分析
名義変数 (生存・死亡)	あり	Cox 比例ハザードモデル

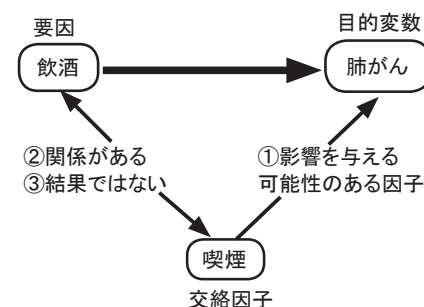


図 1. 交絡因子の条件

ロジスティック回帰分析を行うことにより、目的変数に対して、いくつかある交絡因子がどれほど関連しているのかを推察する。

3) 多重共線性

ロジスティック回帰分析において、共変量間に強い相関があったり、一次従属な変数関係があったりする場合には、解析が不可能であったり、たとえ結果が求められたとしてもその信頼性は低くなったりする。このような場合に共変量間に多重共線性があると言う。共変量間で相関関係をチェックし、相関係数が0.8もしくは0.9を越えるような場合には、多重共線性を疑う必要がある。

また、BMIと体重が同じ回帰式にあった場合、正しい結果が導き出せない。これは、BMIの計算式に体重が入っているため、このような場合に体重とBMIは一次従属な変数関係にあると言い、多重共線性を認める。

多重共線性があるかどうかをチェックする方法としては、各共変量間の相関を調査することや、VIF（分散拡大係数）を統計解析ソフトウェアから求め、 $VIF < 10$ であることを確認する方法がある。多重共線性が認められた場合は、“どちらか一方”を共変量に採用する。

4) サンプル数

ロジスティック回帰分析を行う場合、イベントの少ないほうの発生サンプル数は共変量の数の10倍の例数が必要である。³ 例えば、目的変数が「再発あり・なし」で、共変量が4つあった場合、少ないほう（例えば「再発あり」）のデータが40例以上必要ということになる。

5) モデルの適合度

ロジスティック回帰分析のモデルの適合度を表す指標としては、AIC（赤池の情報量基準）、c統計量（AUC）、寄与率（決定係数）、Hosmer-Lemeshowの適合度等がある。⁷

AICはいくつかのモデルがどの程度現実の観察データにフィットするかを比較する場合の尺度と

なる。その値が小さいものほど良いモデルとみなされる。⁸

c統計量は、ロジスティック回帰の目的変数（0または1）と計算結果である予測確率でROC曲線を書き、曲線下面積（area under the curve; AUC）を求める時に、 $AUC = c$ 統計量となる。c統計量は0.7以上であることが望ましい。⁹

寄与率（決定係数）は、相関係数の2乗で、回帰式によって目的変数が説明できる割合を表わす。ロジスティック回帰分析では、重回帰分析の決定係数に相当するものとして、Nagelkerke R^2 と Cox & shell R^2 という「疑似決定係数」が計算される。0から1の値をとり、1に近いほうが適合度が高い。

Hosmer-Lemeshowの適合度では、有意確率（P値）が0.05より大きければモデルが適合していると考えられる。¹⁰

6) オッズ比と95%信頼区間とP値

ロジスティック回帰分析の結果の解釈については、オッズ比とオッズ比の95%信頼区間およびP値に着目する。おおまかな解釈としては、目的変数に対して、それぞれの共変量が、回帰式にどの程度影響しているのか（オッズ比）、真のオッズ比が高い確率で含まれるのはどの範囲か、その精度はどれくらいか（95%信頼区間）、統計的に回帰式にどれくらい有意に関わっているか（P値）、を示す指標になる。

まず、オッズ比について考えてみたい。臨床研究では、本来、イベントの発生する危険度（リスク）がどれくらいかということを知りたいものである。前向き研究であるコホート研究は、あらかじめ対象者（母集団）を決定しているため、リスク比（相対危険度）を計算することが可能だが、我々が臨床研究でよく用いる後向き研究である症例対照研究（ケースコントロール研究）では、まず症例群（例：再発した患者）が存在し、その後に対照群（例：再発していない患者）を任意に集めるため、確実な母集団を特定できず、リスク比を求めることができない。そこでオッズ比が用いられる。

オッズは、競馬等のギャンブルでよく用いられ、

「勝負は五分五分だ」とか、「四分六だ」という風に使われている。統計学的に表現すると、オッズは事象が起こる確率と起こらない確率の比のことで、いま、ある事象が発生する確率を P とすると、オッズは $P/(1-P)$ になる。さらにこのオッズとオッズの比をオッズ比と呼ぶ（図 2）。オッズに差がなければ、オッズ比は 1 となる。

ある共変量 x_1 が 1 単位変化したときのオッズ比は、回帰係数を a_1 とすると次式で定義される。

オッズ比 $= e^{a_1}$ (e : ネイピア定数)。

ここで注意したいのは、オッズ比が 2 だった場合、オッズの比が 2 倍になったということであり、リスク比が 2 倍になったのではないということである。つまりリスク比とオッズ比は同じものではないことに注意しなければならない（図 3）。ただし、オッズ比は、結果イベントの起こる確率が非常に低い場合（例：有病率が低い場合）は、リスク比に近似可能となる利点もあり、オッズ比が利用されている理由のひとつである（図 4）。

信頼区間とは、推定値の精度の指標である。この区間は、高い確率（通常は 95%）で“真の”値を内包すると考えられるデータに矛盾しない値の範囲を表す。信頼区間は推定値と同じ単位で表現される。広い信頼区間ほど精度が低いことを意味し、狭い信頼区間は精度が高いことを意味する（CONSORT Statement）^{11,12}。オッズ比の 95% 信頼区間を求めた時に、区間の中に 1（オッズ比：1）を含まなければ、そのオッズ比は、有意水準 0.05 で統計学的に有意に回帰式に影響していることを示す。

P 値は、回帰式にどれくらい有意に関わっているかという統計学指標として用いられる。有意水準 5% ならば、 P 値が 0.05 より小さいと有意である。しかしながら、サンプルサイズが大きくなると P 値が小さくなること、再現性が高くないこと、回帰式に対する影響の程度が信頼区間に比べると具体的な数値ではなくわかりづらいことなど、 P 値の使用については、近年様々な議論があり、慎重に用いる必要がある。¹³⁻¹⁷

	疾患あり	疾患なし	全体
曝露あり	a	b	a + b
曝露なし	c	d	c + d

$$\text{リスク比} = \left(\frac{a}{a+b} \right) / \left(\frac{c}{c+d} \right)$$

$$\text{オッズ比} = \left(\frac{a}{b} \right) / \left(\frac{c}{d} \right) = \frac{a \times d}{b \times c}$$

図 2. リスク比とオッズ比の求め方

	疾患あり	疾患なし	全体
曝露あり	100	100	200
曝露なし	50	150	200

$$\text{リスク比} = \left(\frac{100}{200} \right) / \left(\frac{50}{200} \right) = 2$$

$$\text{オッズ比} = \frac{100 \times 150}{50 \times 100} = 3$$

図 3. リスク比とオッズ比が近似しない例

	疾患あり	疾患なし	全体
曝露あり	100	9900	10000
曝露なし	50	9950	10000

$$\text{リスク比} = \left(\frac{100}{10000} \right) / \left(\frac{50}{10000} \right) = 2$$

$$\text{オッズ比} = \frac{100 \times 9950}{50 \times 9900} = 2.01$$

図 4. リスク比とオッズ比が近似する例

7) 論文への報告項目

ロジスティック回帰分析によって得られたさまざまな結果について、どの結果を報告すればいいのだろう。医学論文作成のためのガイドライン第 2 版によると、報告に用いる表には、比較に用いた変数を明記し、各変数を記述統計量で要約する。表には、共変量の回帰係数とその標準誤差、オッズ比、オッズ比の 95% 信頼区間、 P 値を含めるとある。¹⁸ 表 2 は、ロジスティック回帰分析の報告の例である。

表 2. 4 つの共変量があるロジスティック回帰モデルの報告例

因子	係数(B)	標準誤差	Wald χ^2	P 値	オッズ比	95%信頼区間
切片	-2.715	0.651				
X1	0.022	0.008	7.756	0.005	0.98	0.96-0.99
X2	0.806	0.218	13.71	0.001	2.24	1.46-3.43
X3	-0.131	0.137	0.921	0.338	1.14	0.87-1.49
X4	0.041	0.011	15.15	0.001	1.04	1.02-1.06

X2 因子は、目的変数に対して強く影響している (OR = 2.24, 95%CI 1.46-3.43, $P = 0.001$).

3. 傾向スコア (propensity score) 解析

傾向スコア解析は、1983 年に Rosenbaum と Rubin によって初めて提案された.² 複数の共変量を、傾向スコアという一つの数値に集約し、交絡因子の影響を除去する方法として用いられる. 近年、数値のデジタル化に伴い、複雑な計算が容易に処理でき、とくに海外での論文採用数が飛躍的に伸びている.³

1) 傾向スコアとは

傾向スコアとは、ある目的変数があったときに、さまざまな因子から予測された「予測確率」のことである. 例えば、目的変数が「がんの手術をする・しない」で「手術をする」をイベントとした場合に、年齢・性別・転移の有無・腫瘍の大きさ等いくつかの交絡因子から予測される手術が選択される確率のことである. つまり、いくつかの交絡因子を使用して、「手術をするという決定がなされる傾向 (予測確率) はどれくらいなのか」を計算したものになる. 予測確率を推定する方法として、ロジスティック回帰分析やプロビット回帰分析が用いられる.

2) 傾向スコア解析

傾向スコア解析とは、潜在的な交絡要因となる様々な共変量を、ロジスティック回帰分析やプロビット回帰分析で計算した予測確率=傾向スコアという一つの合成変数に縮約 (一次元化) し、その傾向スコアを基準として、交絡因子の影響を除去するためにマッチングや層別化等を行うものである.¹⁹⁻²¹ 以下に傾向スコアを利用した交絡因子の影響を除去する方法を紹介する.

2)-1. 傾向スコアマッチング (propensity score matching)

比較したい 2 群間において、目的変数に対する傾向スコアを算出し、同じ傾向スコア同志の症例をペアとして対応させて、解析するものである. このときのペアの抽出は、キャリパーと言う一定の傾向スコアの幅を持ってマッチングする. したがって、傾向スコアマッチング後の各群の症例数は同じ数になり、以後の統計学的検定は「対応のある 2 群」として処理される.

例えば、あるがんの手術において、手術法の A 法と B 法で、5 年後までにがんの再発があったか・なかったかについての差を比較したいとしよう. この時の目的変数は、再発あり・なしである. 交絡因子として考えられる、年齢・性別・リンパ節転移の有無・癌のステージ・放射線治療の有無・化学療法の有無・喫煙・BMI 等を共変量として、ロジスティック回帰分析を行い、傾向スコアを算出する. A 法と B 法の間で、算出された傾向スコアが同じペアを照合 (マッチング) し、McNemer 検定 (対応のある 5 群の比率の差の検定) を行う (図 5).

実際のデータを用い傾向スコアマッチングにより交絡因子の調整を行った例を呈示する (表 3, 4). 当院で開腹手術を行った 690 例において、輸血施行群と輸血非施行群の 2 群間の違いによって予後が変化するかどうかを調べたいとする. 今回のデータの中で、「輸血施行の意思決定をする際に参考にしたかもしれないデータ」として、手術時間・手術直前のヘマトクリット値・年齢の 3 つの因子を取り上げた. 傾向スコアマッチングを行う前のこれらのデータでは、輸血群と非輸血群で、いずれも有意な差がみられた (表 3).

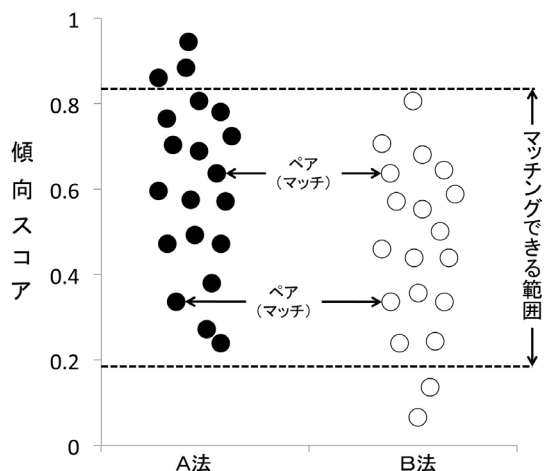


図 5. 傾向スコアマッチング

手術時間・手術直前のヘマトクリット値・年齢を「共変量」とし、輸血あり・なしを「目的変数」としたロジスティック回帰分析を行う。回帰式の結果が、それぞれの症例において、輸血施行の決定がされる確率となり、これが傾向スコアである。例題では、690個の傾向スコアが求められた。輸血あり群と輸血なし群の傾向スコアをマッチングし、傾向スコアの似た者同士のペアを作成する。傾向スコアマッチング後の輸血群と非輸血群で、手術時間・手術直前のヘマトクリット値・年齢のデー

タを比較すると、有意差がみられなくなり（表4）、これらの交絡因子の影響を受けない解析が行うことが可能になった。

ただし、今回の例でもそうであるが、傾向スコアマッチングでは、ペアとなる傾向スコアがない症例、つまり明らかに輸血施行の決定がされる確率の高い症例（傾向スコアが高い）や明らかに輸血をしない確率が高い症例（傾向スコアが低い）は、マッチングの相手が見つからず、対象から除外されるため、マッチするペアが少ない場合、対象数が減少してしまうことがあり、短所のひとつである。

2)-2. 層別解析 (stratified analysis)

傾向スコアの大小により、いくつかの層（サブクラス；通常は5つに分ける）を作成し、層ごとに解析した後、ひとつにまとめて解析する。つまり、傾向スコアの高いものは高いもの同志、傾向スコアの低いものは低いもの同志で比較する（図6）。症例を全例用いて解析することが出来る。クロス集計表から、Mantel-Haenszel 統計量を用いて判定する。

ただし、層別解析では、各群において、傾向スコアが反対側に偏ると、傾向スコア小の層と大の層で、どちらかの例数のバランスがとれなくなる場合があるので、確認が必要である。

表 3. 開腹手術における輸血施行群と輸血非施行群の手術時間・手術直前のヘマトクリット値・年齢

	輸血施行群（平均±SD） N = 341	輸血非施行群（平均±SD） N = 349	P値
手術時間（時間）	4.67 (2.08)	2.33 (1.06)	<0.001
術前ヘマトクリット値（%）	36.3 (9.0)	42.5 (14.2)	<0.001
年齢（歳）	56.5 (13.5)	52.8 (14.3)	0.001

表 4. マッチング後の開腹手術における輸血施行群と輸血非施行群の手術時間・手術直前のヘマトクリット値・年齢

	輸血施行群（平均±SD） N = 138	輸血非施行群（平均±SD） N = 138	P値
手術時間（時間）	3.11 (1.19)	3.04 (1.08)	0.606
術前ヘマトクリット値（%）	37.0 (11.2)	38.9 (8.0)	0.114
年齢（歳）	56.2 (14.4)	54.2 (14.7)	0.258

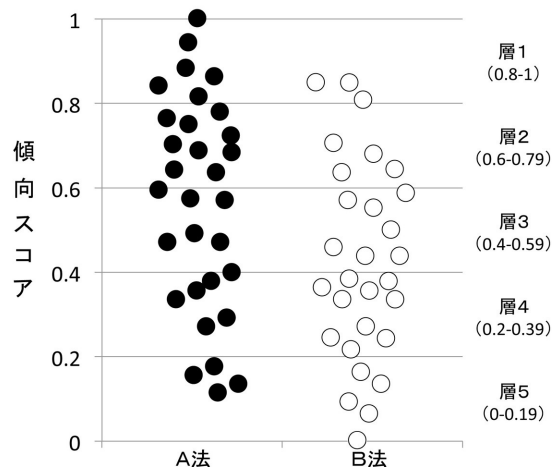


図 6. 層別解析

2)-3. 共変量として用いる

傾向スコアを年齢や性別などの交絡因子を1つにまとめた代表と考え、傾向スコアを共変量として回帰分析を行う。

2)-4. IPW (inverse probability weighting) 法：逆確率重みづけ法

傾向スコア（予測確率）の逆数を用いて、共変量に重みをつけて解析を行う方法である。傾向スコアマッチングの際に、マッチするペアが見つからず、対象から除外した症例を解析に含めることができる。

目的変数の「イベントあり」の群の重みは傾向スコアの逆数 ($1/P$) を用い、「イベントなし」の群の重みは1から傾向スコアを引いた数（イベントでない確率 $1/(1-P)$ ）を用いる。群間比較の際には一般化推定方程式を使用する。

さらに近年、拡大された IPW 推定量とし、結果変数の周辺期待値や因果効果を正しく推定できるモデルとして、「二重にロバストな推定量 (doubly robust estimator) 法」等が提案されており²²⁻²⁴ さらなる発展をみせている。

4. おわりに

ロジスティック回帰分析と傾向スコア分析について、概要を解説した。

臨床研究において、とくに後ろ向き研究においては、発症や死亡等の目的変数と、治療の効果の判定や曝露因子による影響等の関係を調査するときに、対象における偏り（バイアス）や交絡因子の影響は必ず考慮に入れなければならない。

研究計画を立て、目的変数に関わるさまざまなデータを収集した後、解析を行う際には、まず、それぞれの項目の平均や標準偏差・カイ2乗検定・平均の差の比較等の単変量解析を行い、背景因子や交絡因子を把握することが重要である。

単変量解析の結果をふまえたうえで、多変量解析であるロジスティック回帰分析を行い、目的変数に対するそれぞれの交絡因子がどれだけ影響しているのかを推察する。

また、交絡因子をひとつにまとめた傾向スコアを利用して、交絡因子の影響を除去する方法が提案され、傾向スコアを利用した論文投稿も増加している。傾向スコアの用法にはいくつかの方法があり、対象や交絡因子により方法を選択する。また、傾向スコアの発展した使用法が提案されており、傾向スコア解析法は、今後のさらなる展開が注目される。

参考文献

1. 佐藤俊哉. 疫学研究における交絡と効果の修飾. 統計数理 1994;42:83-101.
2. Rosenbaum PR, Rubin DB. The central role of the propensity score in observational studies for causal effects. Biometrika 1983;70:41-55.
3. 新谷 歩. 多変量解析－説明変数の選び方. 今日から使える医療統計. 東京:医学書院;2015:101-113.
4. 奥田千恵子. 論文に多用される多変量解析法・1 総論. 総合リハビリテーション 2015;43:547-552.
5. 山岡和枝. 多変量解析とは何か. Diabetes Frontier 2015;25: 603-609.
6. 中村好一. 基礎から学ぶ楽しい保健統計. その他の多変量解析・14. 公衆衛生 2015;79:796-800.
7. 瀧口 徹. 歯科疫学統計 (第3報) 重回帰分析, 多重ロジスティック回帰分析モデルの適合度判定指標の解釈. SPSS, STATA の利用に際して. ヘルスサイエンス・ヘルスケア 2005;5:35-49.
8. 宮副一郎. ロジスティック回帰分析入門 (3) 出力結

- 果の読み方. 日本保険医学会誌 2009;107: 215-227.
9. 久繁哲徳. 検査の有用性の指標と ROC 分析. 臨床検査 2005;49:1335-1340.
 10. Hosmer D W, Lemeshow S. Applied Logistic Regression Third edition (Wiley Series in Probability and Statistics). Wiley.;2013.
 11. CDISC (Clinical Data Interchange Standards Consortium) 標準推進プロジェクト, CDISC Clinical Research Glossary 信頼区間, (http://www.trikobe.org/cdisc/glossary/glossary.php?mode=detail&id=uid_110) Accessed 2016 Feb 18.
 12. Kenneth F Schulz, Douglas G Altman, David Moher, for the CONSORT Group 訳. 津谷喜一郎, 元雄良治, 中山 健夫. CONSORT 2010 声明 ランダム化並行群間比較試験報告のための最新版ガイドライン. 薬理と治療 2010;38:939-947.
 13. 前谷俊三. 医療におけるエビデンスと P 値. 天理医学紀要 2003;6:54-79.
 14. Sackett DL, Straus SE, Richardson WS, et al. Evidence-based medicine: How to practice and teach EBM. 2nd Ed, Churchill Livingstone, 1988.
 15. Nuzzo R, Scientific method: statistical errors. Nature 2014;506:150-152.
 16. Statement on Statistical Significance and P-Values March 7, 2016. (<http://www.amstat.org/newsroom/pressreleases/P-ValueStatement.pdf>). Accessed 2016 Mar 14.
 17. Ronald L. Wasserstein & Nicole A. Lazar. The ASA's statement on p-values: context, process, and purpose, The American Statistician. (<http://amstat.tandfonline.com/doi/suppl/10.1080/00031305.2016.1154108>). Accessed 2016 Mar 14.
 18. 大橋靖雄, 林健一 (監訳). わかりやすい医学統計の報告 医学論文作成のためのガイドライン第2版. 東京: 中山書店 ;2011
 19. 星野崇宏・繁樹算男. 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 2004;31:43-61.
 20. 星野崇宏・岡田謙介. 傾向スコアを用いた共変量調整による因果効果の推定と臨床医学・疫学・薬学・公衆衛生分野での応用について. 保健医療科学 2006;55:230-243.
 21. 星野崇宏. 調査観察データの統計科学. 東京: 岩波書店 ;2009.
 22. Bang H, Robins JM. Doubly robust estimation in missing data and causal inference models, Biometrics 2005;61: 962-972.
 23. Rotnitzky A, Robins J. Analysis of semi-parametric regression models with non-ignorable non-response. Stat Med. 1997;16:81-102.
 24. 逸見昌之. 欠測データに対するセミパラメトリックな解析法 - その理論的背景について - 統計数理 2014;62:103-122.

Interpretation of logistic regression and propensity score analysis

Hitoshi Obayashi

Tenri Institute of Medical Research

One of the multivariate analyses commonly used in medical research is logistic regression. It is designed to predict from explanatory variables (covariates) the probability of binary (nominal) outcomes (P), such as survival or death and positive or negative response to a test drug, using a logistic regression model, where $\log(P/(1-P))$ is assumed to be a linear combination of covariates ($b_0 + b_1x_1 + b_2x_2 \dots$).

In non-randomized retrospective studies, the results of these analyses may be biased by confounding factors. To adjust for the confounders, a propensity score analysis using logistic regression has been proposed, permitting us to abridge multiple confounders to a single composite variable. This concept and technique are herein explained.

Keywords: logistic regression, propensity score, confounder