原稿作成日：　2017 年 3 月 31 日

検定の選び方：

検定は結論を変え得る！

不適切な検定を故意に選ぶのは不正行為

〈教材提供〉
AMED 支援「国際誌プロジェクト」　提供

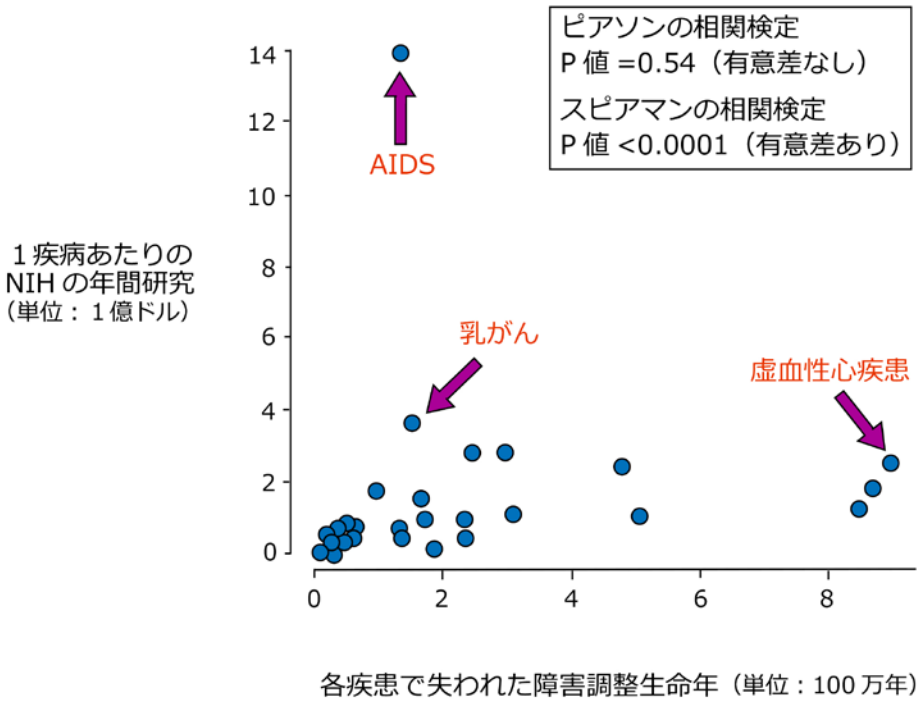
草案
新谷歩　大阪市立大学医学研究科医療統計学講座教授
加葉田大志朗　大阪市立大学医学研究科医療統計学講座特任助教

査読
大門貴志　兵庫医科大学医療統計学教授
角間辰之　久留米大学バイオ統計センター教授
市川家國　信州大学特任教授
山本紘司　大阪市立大学大学院医学研究科医療統計学講座准教授
石原拓磨　大阪市立大学大学院医学研究科医療統計学講座特任助教

目次

はじめに

下の図は、アメリカ国立衛生研究所（National Institutes of Health; NIH）のある年の疾患ごとの研究費を縦軸に、横軸にはそれぞれの疾患によって失われた年数（障害調整生命年）をプロットした散布図です。2 つの変数に相関があることが示されれば、米国 NIH は国民の命を救うために研究費を有効に活用していることになります。2 つの連続変数の相関を調べる際によく使われる検定に、ピアソンの相関係数に対する検定とスピアマンの相関係数に対する検定があります。ここでピアソンの相関係数に対する検定の P 値は 0.54 であり、一方、スピアマンの相関係数に対する検定では P 値は<0.0001 とほぼ 0 に近い値となりました。このように相関係数に対する検定であっても、どちらの検定を利用するかによって結果が大きく異なることが分かります。



お粗末な医学研究がもたらすスキャンダル

イギリスの著名な統計家であるダグラス アルトマン教授は「Medical Scandal」と題して、以下のように言及しています。『故意にまたは知らず知らずに間違った治療を行ったり、正しい治療を間違った方法で行う（投薬量を間違えるというような）医師のことをどう思いますか？たいていの人は、そのような行動はプロフェッショナルでない、間違いなく非倫理的だ、あるいは全くもって許しがたいと考えるでしょう。では、誤った検定手法を故意に、あるいは知らず知らずに使う、もしくは正しい検定手法を誤った方法で使用する、そして検定結果を間違って理解して発表する、といったことによって間違った結論を導く研究者の場合はどうでしょうか？実は、数え切れないほど多くの一般的に加え専門的な医学研究論文でこのようなことが日常的に行われていると指摘されています。これは間違いなくスキャンダルというものです。』（DG Altman, The scandal of poor medical research *BMJ* 1994;308:283）

基礎研究で多用されるスチューデントのｔ検定のほかに、ウィルコクソン順位和検定や、カイ二乗検定、ログランク検定など、多くの検定が国際誌中の論文では用いられています。「自分の行っている検定が正しいのかわからない」、「いろいろな検定があるけれど、検定の選び方によって結果が全く違ってしまう」などという悩みを持っている方も多くいるのではないでしょうか。アルトマンが指摘しているように、誤った解析で誤った結論を導いてしまうと、再現性が失われ、科学的に問題があるというだけでなく、効果があるものを効果がないと結論づけてしまったり、効果がないものをさも効果があるように結論づけてしまうということさえあるのです。研究結果を待ち望んでいる多くの患者さんや社会へ間違った結果を伝えてしまうことになり、まさに非倫理的なのです。ましてや自分の都合の良い結果が出る解析だけを意図的に利用するのは、研究不正とみなされるもので、研究者自身のしっかりとした倫理感が必要です。以下では検定の誤用を防ぐために、正しい検定の選び方について解説しました。

学習目標

本単元を通じてあなたが修得を目指すものは：

- 検定の選び方の基本原則を知る
- データの分布をチェックすることができる

検定の選び方

検定を選ぶ際に必要な情報

効果の推定やＰ値を計算するための検定には様々なものがあります。同じデータでも、異なる検定を用いれば異なるＰ値が計算されます。上記のNIHのデータの例のように、ピ

アソンの相関係数に対する検定とスピアマンの相関係数に対する検定の両方を行って、有意差の出た検定の結果を論文に記載するなど、手当たり次第検定を繰り返して、研究者にとって最も望ましい（多くの場合はＰ値が小さくなる）結果を選ぶというのは科学的にも倫理的にも妥当ではありません。そのため、研究の最終解析で用いる検定手法は、あらかじめ研究実施計画書に記載しておく必要があります。国際誌のチェックリストでは検定が適切に選択されているかどうか、また最終解析にどの検定を用いるか計画段階から固定されていたかどうかについての項目があります。どの検定を用いるかは、どのようなデータでどのような比較を行うのかによって異なります。それではどのような手順で適切な検定を選べばよいのでしょうか。その判断に必要な項目は以下のとおりです。これらは、統計家に相談した際に、彼らにとっての関心事になります。各々を順番にみていきましょう。

項目①　無作為化が行われたか？
項目②　着目するのは差か相関か？
項目③　対応のあるデータか？
項目④　アウトカムの変数の種類は？
項目⑤　アウトカムが連続変数の場合、正規分布に従っていると仮定できるか？
項目⑥　比較群の数は？
項目⑦　症例数は？

項目①　無作為化が行われたか？

ある治療法の効果を調べるときに、この治療法を用いた群とそうでない群に単純に分けて死亡率などのアウトカムを評価するとき、治療された人の方の群に具合の悪い人が多く含まれていることもありうるので、治療された患者さんの死亡率が治療によって軽減されなかったという結果が出ても本当にこの治療が効かないのか、もともと治療群に具合の悪そうな人が多く含まれていたからなのかわかりません。このように比較グループの背景に関して群間に差が生じてしまうと、治療の効果を正しく解析することができなくなります。

無作為化を行わない研究の比較群間の背景

背景因子	アスピリン使用群 (N=2310)	アスピリン非使用群 (N=3864)
年齢（平均）	62 歳	56 歳
男性 (%)	77%	56%
糖尿病歴 (%)	17%	11%
高血圧 (%)	53%	41%
冠動脈疾患歴 (%)	70%	20%

上の表のような比較群間の背景の偏りを防ぐために、通常行われるのが無作為化です。コインを投げて表が出たら治療あり群、そうでなければ治療なし群に、といった具合に無作為にどちらかの群に割り付けていくことを無作為化と称しています。このように誰が治療あり群で誰が治療なし群かを無作為に割り付ける研究では、通常、下の表のように、治療の有無以外の背景情報は群間で揃う傾向にあるため、比較群間の背景の違いを解析で考慮する必要は一般的にはありません。

無作為化を行った研究の比較群間

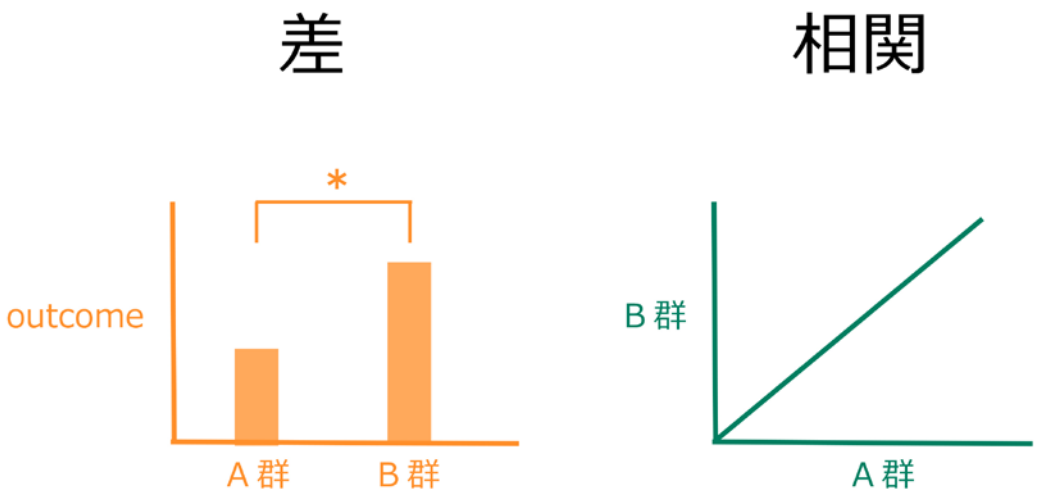
背景因子	アスピリン使用群 (N=2226)	アスピリン非使用群 (N=2269)
年齢（平均）	64 歳	64 歳
男性（%）	43%	42%
高血圧 (%)	69%	68%
高脂血症 (%)	41%	36%
肥満 (%)	22%	24%
糖尿病 (%)	17%	16%

そのため、無作為化が行われた研究では、アウトカムを治療あり群と治療なし群で直接比較する検定を行うことが可能です。このような解析を「単変量解析（Uni-variable analysis)」と呼びます。

無作為化が行われない研究の場合、比較群間の背景が揃わないことが多いので、その場

合は多変量解析（Multivariable analysis）の統計手法を用いて背景の違いを調整することが一般的です。

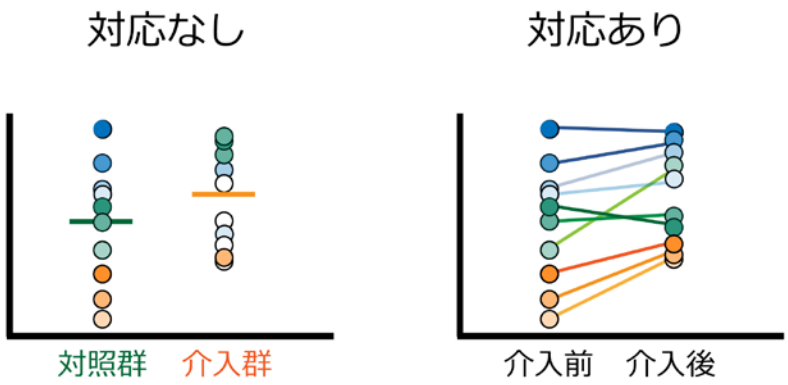
項目②　着目するのは差か、相関か？



例えば、降圧薬 A を服用した群（介入群）としなかった群（対照群）を比較する研究において、アウトカムとしての研究後の平均血圧を比べるような、2 群間のアウトカムの差を比較する場合は、スチューデントの t 検定を用います。そうではなく、1 群のデータでコレステロール値が上がれば血圧も上がる、というような２つの変数間の関係の有無を見たい時には、相関に関する検定を用います。

項目③　対応のあるデータか？

上記の、差に着目する検定を用いる場合には、介入群と対照群のデータが独立、つまり両群間で無関係な研究対象者によるデータであるかどうかによっても、検定手法が変わります。



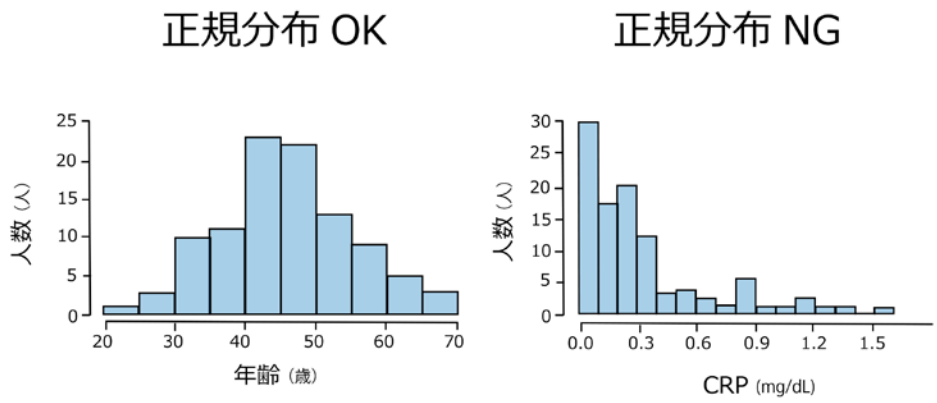
全く別々の研究対象者が介入群と対照群に入っているのであれば、両群のデータは独立である、つまり対応していないとみなし、この場合は上述したスチューデントのｔ検定などを使うことができます。一方、研究対象者に先ず既存薬を服用してもらってそのアウトカムを評価した後（介入前）、（既存薬の効果が消失するだけの）十分な期間をおいて、次に新薬を服用してもらって後、そのアウトカムの評価を行う（介入後）といった、研究があります。このように同じ研究対象者のデータを繰り返し収集するような場合には、介入前と介入後の 2 つのデータ間には対応があるとみなします。このようなケースでは両群のデータ間に対応があるため、上記とは異なる検定を用いる必要があります。例えばスチューデントのｔ検定ではそれぞれの群の平均値を計算してその差を比較していましたが、対応のあるｔ検定ではそれぞれの研究対象者の中でのアウトカムの変化量を計算して、変化量の平均が 0 かどうか（つまり、差がないか否か）の仮説を立てて検定を行います。ｔ検定以外の検定においても、一人から 1 回しかとっていないデータを独立の複数群で比べるのか、同じ人から何回も繰り返し収集しているデータを比べるのかによって、用いるべき検定手法は異なりますので、注意が必要です。

項目④　アウトカムの変数の種類は？

アウトカムの変数の種類によっても検定手法は異なります。データは大きく「連続変数」と「カテゴリカル変数」の 2 つに分けられます。年齢や血圧など数字が連続しているようなデータを「連続変数」と称します。それ以外のデータをカテゴリカル変数と呼びますが、その中でも病気の有無や性別など 2 種類に分けることができるデータは「2 値変数」と呼ばれます。病気の重症度など、度合いに意味を持つデータは「順序のあるカテゴリカル変数」、動物種や人名など、度合いなどを持たないデータは「順序のないカテゴリカル変数」と呼ばれています。先ほどまで例に出していたスチューデントのｔ検定は、アウトカムが連続変数であるときに使うことができます。

連続変数	カテゴリー変数
年齢	順序変数 がんのステージ
血圧	名義変数 併用薬剤
BMI　など	2 値変数 生存 / 死亡　など

項目⑤　アウトカムが連続変数の場合は正規分布に従っていると仮定できるか？



アウトカムの変数が連続変数の場合、そのデータが正規分布に従っていると仮定できるかどうかにも注意が必要です。先ほど紹介したスチューデントのｔ検定は、アウトカムの平均値を独立な群間で比べる検定です。平均値や標準偏差は、データが正規分布に従うと仮定できるときに正しくデータを記述することができます。そして、それらの平均値や標準偏差を用いる検定は正しい結論を導きます。このように正規分布など、何かしらの分布パターンを想定したときにパフォーマンスが良くなる解析を「パラメトリック検定」と呼びます。（正確にはスチューデントのｔ検定では、各群が独立であること、各群のアウトカムの変数が正規分布に従っていると仮定できること、各群の標準偏差が似通っていること、という仮定を満たすときに最も検出力が高くなると考えられています。）
その反対に正規分布などの分布を想定しない解析をノンパラメトリック検定と呼びます。このような検定ではデータの順位（ランク）を使うことが一般的です。アウトカム変数が正規分布に従っていると仮定できるか否かは、箱ひげ図やヒストグラムなどで分布の形状を目視する、または正規性の検定（例えば、コルモゴロフ-スミルノフ検定）を行うことで確認することができます。

項目⑥　比較群の数は？

比較群の数によっても、検定の選び方に異なりがあります。単元 11 では、3 つ以上の比較群がある場合には複数の検定を行う必要があるため、多重性の問題に対する、対処策が必要であることを説明しました。その対処策の 1 つにグローバル検定を用いて、全ての群の中でどれか 1 群でも異なる群があるかどうかの検定を行うことを説明しました。

例えば、「低用量、中用量、高用量の薬剤治療をうけた 3 つの群で投薬後の平均血圧を比較する」という研究を考えてみましょう。用量の低中高の順序をとくに考慮する必要がなければ、この研究で得られたデータついてグローバル検定としての役割を果たす分散分析を適用することができます。そこでは、3 つの群の平均血圧が同時に全て等しいという帰無仮説の是非について検定を行うことになります。この検定では計算される P 値は一つだけであり、多重性の問題が生じることはありません。一方、比較群が 2 群の場

合、P 値が一つ計算され、そして複数群の場合には、複数の P 値が計算されてきます。この場合は、多重性の問題に対処する必要が生じて P 値の補正が必要になります。しかし、前記のグローバル検定であれば、P 値は一つしか計算されないのでこの補正が必要なくなります。そのため、比較群の数も検定を選ぶ上での重要な要素になるのです。

項目⑦　症例数は？

病気や生存などのイベントの有無を示す 2 値変数をアウトカムとして扱う時には、その発生割合が異なるかどうかを検定します。その際、総症例数が 40 例以上の場合にはカイ二乗検定、20 例未満の場合にはフィッシャーの正確検定を用いることが勧められています。また、総症例数が 20～39 例の場合には、比較群それぞれのイベントありとなしの数の「期待値」によってどちらを使うかを決定することがあります。期待値とは「介入群と対照群で違いがない」と仮定した場合の、介入群および対照群それぞれにおけるイベントあり症例とイベントなしの症例数のことを指します。

例えば総症例数が 30 例だった場合に介入群と対照群にそれぞれ 15 人の研究対象者がいるとします（下表）。この時に全体では 10 人の研究対象者にイベントが起こった、つまり全体ではイベントの割合が 3 分の 1 （33%）だとします。この場合の期待値とは実際に観察されたデータ（左表）ではなく、治療法とイベントの間に全く関連がない時に期待されるデータ（右表）を意味します。つまりもし介入群と対照群でイベントの占める割合が同じであると考えた場合には、介入群では 15x1/3 =5 人がイベントありの症例、残りの 10 人がイベントなしの症例、また対照群においては 15x1/3=5 人がイベントありの症例、残りの 10 人がイベントなし症例という数字になると期待されます。実際には対照群の 7 人に比べて介入群の方がイベント数が少なく 3 人という結果でした。この場合、それぞれの群でイベントありの症例、あるいはなしの症例の期待人数が 5 人以上になりますので、このケースではカイ二乗検定を用いることになります。

観察されたデータ			
	介入群	対照群	計
イベントあり	3	7	10
イベントなし	12	8	20
計	15	15	

関連がない場合に期待される結果			
	介入群	対照群	計
イベントあり	5	5	10
イベントなし	10	10	20
計	15	15	

これまでに挙げた 7 つの項目の組み合わせから、適切な検定を選ぶことになります。以下の表に検定の選び方に関する項目と検定の名称をまとめています。参考にしてください。この表の使い方はビデオでも説明しています。各検定の詳細については次の単元で説明しますが、統計ソフトウェアを使っていく上での手順についてもビデオを参照してください。

Q1. ランダム化	Q2. 差か 相関か？	Q3. データ間 の対応	Q4. アウトカム 変数の種類	Q5. アウトカム 連続変数の 正規性	Q6. 比較群数	Q7. 症例数	適切な統計テスト
あり	差	対応なし	連続	あり	2		スチューデントのＴ検定
					3 以上		分散分析 (ANOVA)
				なし	2		マン・ホイットニーのＵ検定／ウィルコクスンの順位和検定
					3 以上		クラスカルワリス検定
			カテゴリー		制限なし	総数 20 未満	フィッシャーの正確検定 *
						総数 20 以上	ピアソンのカイ 2 乗検定 *
			2 値打ち切り		制限なし		ログランク検定
		対応あり	連続	あり	2		対応のあるＴ検定
					3 以上		反復データによる ANOVA
				なし	2		ウィルコクスンの符号付順位和検定
					3 以上		フリードマン検定
			2 値		2		マクネマー検定
	相関		連続	あり			ピアソンの相関関係
				なし			スピアマンの相関関係
			2 値				ケンドールの相関関係
なし	重回帰分析による交絡の補正						

新谷歩「みんなの医療統計 12 日間で基礎理論と EZR を完全マスター!」講談社から引用

2 群間の平均を比較する：スチューデントの t 検定とウィルコクソンの順位和検定

ここでは最も広く利用されているスチューデントの t 検定から説明していきます。t 検定は 2 群の母平均の差を比較する検定です。母平均とはデータが集められたもとになる集団を母集団と呼び、その母集団における平均のことを指します。最も広く利用されている検定ですが、この検定によって P 値を正しく計算するには、データが以下の 3 つの仮定を満たしていなければなりません。

スチューデントの t 検定における仮定

① データの独立性：データ間に相関がないこと
② データの正規性：アウトカムに対するデータが、比較する 2 群のそれぞれで正規分布に従っていること
③ データの等分散性：アウトカムに対するデータのばらつきが両群間で等しいこと

① のデータの独立性は、エクセルなどに入力したデータが一人 1 行で入っているか確認してください。一人の研究対象者のデータが 2 行以上に入ってきた場合は、スチューデントの t 検定は使用することはできません。
② のアウトカムデータが各群それぞれにおいて正規分布に従っているか否かは、コルモゴロフ-スミルノフ検定などの正規性の検定で確認することも可能ですが、これは参考程度にした方がよいでしょう。というのは、検定の P 値は、一般に、効果量が一定の場合、標本サイズが大きくなるにつれて小さくなるためです。正規性

の検定でも同様に、データが正規分布に従っていても、その標本サイズが大きくなると、P 値が小さくなり、正規分布とは統計的に有意に差がある（これを正規性が成り立っていないなどと表現します）と判断してしまいます。逆に、データが正規分布に従っていなくても、その標本サイズが小さくなると、P 値が大きくなり、正規分布とは統計的に有意に差があるとはいえない（これを正規性が成り立っていないとはいえないなどと表現します）と判断してしまいます。また、箱ひげ図やヒストグラムなどによる分布を目視で確認する方法もありますが、標本サイズが小さいときにはその確認が困難な場面もあります。この難点を意識したうえで、これらコルモゴロフ-スミルノフ検定などの正規性の検定を補助的に利用することになります。過去の研究結果から、あるいはデータの特性を意識して、データがどのような分布に従っていたかを確認することも役立ちます。

③ のアウトカムデータのばらつきが等しいか否かについては「F 検定」を利用することができます。この F 検定の P 値が 0.05 未満であれば「2 群の分散が異なる」という結論になり、反対に P 値が 0.05 以上であれば「2 群の分散が異なるとは言えない」ということで、慣習的に等分散性が満たされているとみなします。もし等分散性が満たされていない場合には、スチューデントの t 検定ではなく、ウェルチの t 検定という検定を用いることになります。これは EZR というソフトウェアでは「等分散性が満たされていない」という項目にチェックした場合に用いる検定です。F 検定についても正規性の検定と同様に、症例数の数が少ない場合に P 値のみで判断を下してしまうのは適切でない場合があるため、注意が必要です。この難点をも鑑みて、等分散の仮定が満たされているかどうかによらず、ウェルチの t 検定を利用することが勧められることがあります。

さて、アウトカムの変数の分布が正規分布に従っていないと考えられる時はどうするのでしょうか？いくつかの解決法はありますが、まず考えられるのが、正規分布を仮定しないノンパラメトリック検定です。上述のとおり、検定にはスチューデントの t 検定のようなパラメトリック検定とウィルコクソンの順位和検定（マンホイットニー U 検定）のようなノンパラメトリック検定の２通りがあります。パラメトリック検定とはデータに何かしらのパラメータをもつ分布を仮定する検定であり、ノンパラメトリック検定とはそのような分布を仮定しない検定です。アウトカム変数の分布が正規分布に従っていないと考えられる場合は、スチューデントの t 検定の代わりにウィルコクソンの順位和検定を用い得ます。

ノンパラメトリック検定は、データの順位に注目していることが多く、データの分布が正規分布に従っていると仮定できるという条件を課していません。パラメトリック検定の多くにはそれに対応するノンパラメトリック検定があります。アウトカムの分布によって使い分ける事もひとつですが、最近では分布によらずノンパラメトリックな検定の利用が勧められることもあります。国際誌のチェックリストでは、利用した検定の仮定が十分に満たされているかどうかという情報の提供を推奨する傾向がありますので、パラメトリック検定を利用する場合には、データの分布が諸種の仮定を満たしているかどうかを記載する必要があります。

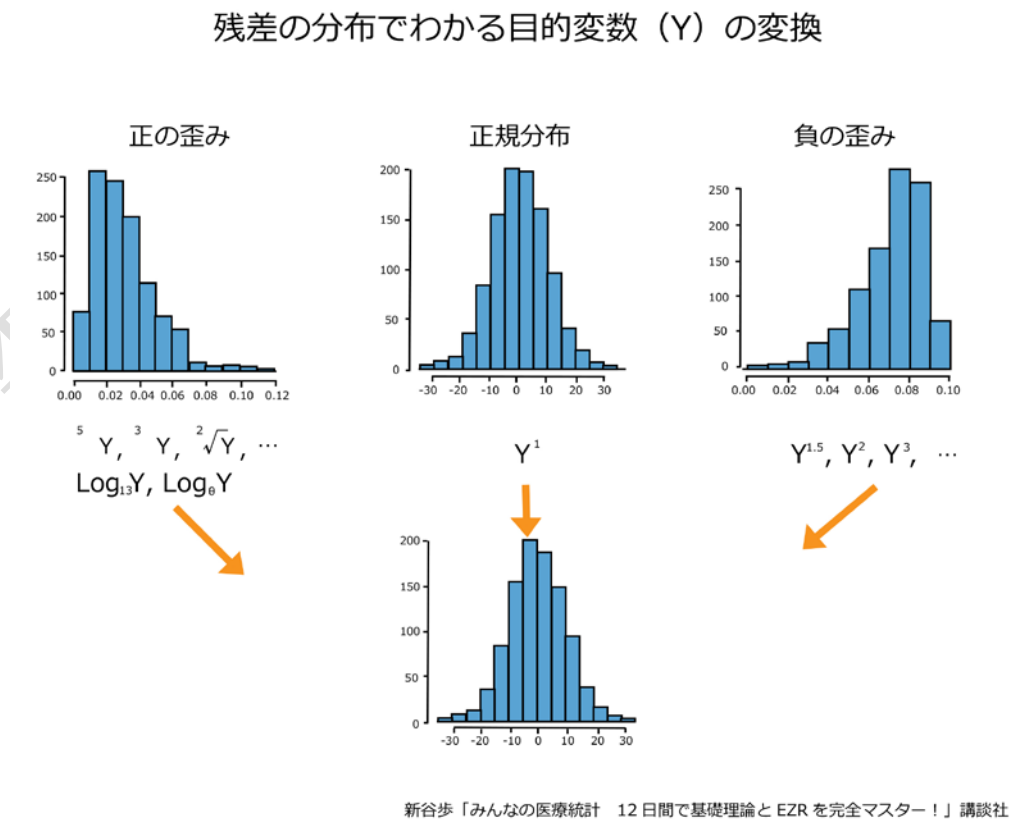
歪んだ分布

歪んだ分布の場合

データが正規分布に従っていると仮定できない場合、ノンパラメトリック検定を用いるという方法のほかに、データを変換して正規分布に従うと仮定できるようにしてから検定する方法があります。単変量解析ではノンパラメトリック検定を選択することはできますが、後の単元で説明する多変量解析では、そのノンパラメトリック検定のような標準的な選択肢は基本的にはありませんので、このデータ変換の方法が役立ちます。

データの変換

医学研究においてよく利用される血清マーカーは、ほとんどのデータが 0 に近い値を取りますが、とかく極端に高い値も存在する歪んだデータとなるものです。このような右側に裾を引いている分布に見られる歪み方を「正の歪み」と呼びます。このようなケースでは、データに対して対数変換などを行うと、分布が正規分布に近づくようになります。この点、対数変換がたまたまよく利用される変換ということだけであって、二乗根や三乗根で変換をもちることもあります。またその反対に左側に裾が長い分布は「負の歪み」を持つ分布と呼ばれます。負の歪みを持つ場合にも、データを二乗や三乗することで正規分布に近づけることができます。



新谷歩「みんなの医療統計　12 日間で基礎理論と EZR を完全マスター！」講談社

どのようなデータ変換にするかを選択するに当たって、スチューデントのｔ検定などの検定を行って､その結果であるＰ値が最も小さくなるものを選ぶようなことは不適切です。最終的な検定を行う前に、データが正規分布に従っているかどうかを上述した方法で確認し、最終的に最も適していると考えられるデータ変換方法を選択した上で、検定を行なってください。

対応のある検定

対応のあるｔ検定

先ほどの検定の選び方の項目③の「対応のあるデータか？」について、各群に同じ研究対象者のデータが含まれる場合などには、データの対応を考慮した検定が必要になることを説明しました｡ここでは対応のある場合のｔ検定について説明していきたいと思います。

先ほど簡単に触れたように、対応のあるｔ検定ではまず、両群内で一人一人の研究対象者についての差分を計算して､その平均値を計算します｡この平均値が０から離れている、つまり同じ症例の間の差が０かどうかの検定をすることになります。この検定では比較するのは同じ研究対象者同士なので、背景の偏りなどのバイアスが入り余地はありません。また同じ研究対象者の中での差分をとるため、データ間のばらつきが小さくなることが多いので、対応のある検定というのは２群間の差の検出力が高くなる傾向があります。

またこれもｔ検定の一種ですので、データが正規分布に従っているという仮定のもとで最も検出力が高くなるパラメトリック検定になります。ここで必要となる前提条件は「研究対象者の中でのアウトカムの変化量が正規分布に従っている」ということであり、この点で、スチューデントのｔ検定とは異なります。もしこの仮定が満たされていない場合には、スチューデントのｔ検定同様に、ノンパラメトリックな検定を利用することが考えられます。

ウィルコクソンの符号付き順位和検定

対応のあるｔ検定に対応するノンパラメトリックな検定は、ウィルコクソンの符号付き順位和検定になります。例えば、介入治療の使用の前後での血圧の平均値を比較するとします。これは研究対象者内で計算した血圧の変化量がプラスだった人（介入後の血圧の方が高い人）と変化量がマイナスだった人（介入後の血圧の方が低い人）の２群で症例数での重みを考慮した変化量の順位の絶対値の和（考え方としては平均値と同じ）を比べることになることから､「符号付順位和」という名前がついています。

	ID	使用前	使用後	差	符号	差の絶対値	順位
	1	110	111	1	＋	1	1
	2	115	118	3	＋	3	2
	3	116	124	8	＋	8	7
	4	95	104	9	＋	9	8
	5	90	103	13	＋	13	9
	6	130	126	-4	―	4	3.5
	7	122	126	4	＋	4	3.5
	8	115	110	-5	―	5	5
	9	98	105	7	＋	7	6
	10	105	120	15	＋	15	10
平均		109.6	114.7	5.1			
標準偏差		12.6	9.2	6.6			

新谷歩「みんなの医療統計 12 日間で基礎理論と EZR を完全マスター!」講談社から引用

この単元に関係するビデオ教材

t 検定
t 検定の仮定
データ変換と t 検定
マンホイットニーの U 検定
対応のある t 検定
対応のある t 検定の仮定
ウィルコクソン符号付順位検定
分散分析（ANOVA）
クラスカルワリス検定
相関の検定

本単元は日本医療研究開発機構:研究公正高度化モデルである「医系国際誌が規範とする研究の信頼性にかかる倫理教育プログラム」（略称：AMED 国際誌プロジェクト）によって作成された教材です。作成および査読等に参加した専門家の方々の氏名は、冒頭に掲載されています。

この単元に関係する国際誌におけるチェックポイントをいくつか紹介します。
（内容は解釈を助けるために一部意訳している部分もあります）

①Nature
（http://image.sciencenet.cn/olddata/kexue.com.cn/upload/blog/file/2010/12/2010128212513557501.pdf; visited on 2018.02.11）
②New England Journal of Medicine （http://www.nejm.org/page/author-center/manuscript-submission#electronic; visited on 2018.02.11）
③Science （http://www.sciencemag.org/authors/science-editorial-policies; visited 2018.02.11）
④The EMBO Journal （http://emboj.embopress.org/authorguide#embargopolicy; visited on 2018.02.11）
⑤JAMA （http://jamanetwork.com/journals/jama/pages/instructions-for-authors; visited on 2018.02.11）

①Nature

- 比較している群（対称群）を明記すること
- 仮説検定の名称を明記すること
- 全ての統計手法を誤解のないように記述すること
- 適用した手法が適切なことを記載すること
- 使用されたデータが用いられた検定手法の仮定を満たしていること（例：分布に歪みがないか、症例数が少なくないかなど）
- 一般的ではない、あるいは複雑な解析手法については Nature の幅広いジャンルの読者が理解できるように詳細を記載すること（もし説明が長くなる場合は補足資料を準備すること）
- 対数（log）変換などデータの変換を行なった場合にはその方法と理由を明記すること

②New England Journal of Medicine

- 3 群以上の複数群間でそれぞれの 2 群ごとにアウトカムを比較する場合、多重検定による第 1 種の過誤の補正を考慮に入れた検定手法を用いるべきである。正規分布にデータが従う場合、t 検定は正しく、また十分にデータの数が大きくなれば t 検定の結果は頑健である。データ数が小さく、先行文献においてデータの歪みが示唆されている場合は、ノンパラメトリック検定を適宜用いるべきである。

③Science

- 十分に知識のある読者であれば、論文作成に使用されたデータを利用して結果を検証することが可能となるように、統計解析の詳細を記載しなければならない。
- t 検定や、ウィルコクソン符号付順位和検定、回帰係数についての Wald 検定など、連続変数の解析に用いられる検定や、正規分布を仮定した 95%の信頼区間、平均と 2×標準偏差、尤度比を用いた信頼区間など、検定に何を用いたのかなどを Materials と Methods の章に明記すること。図の凡例中に解析に用いた検定を実験ごとに記載すること。
- 使用した検定において、その正当性を読者が評価できるよう、それぞれの検定が前提とする仮定・条件が検証されたかについても記載すること。（例）データが正規分布に従っている、生存解析に用いられたデータは比例ハザード性を有するなど。

■ 新たな手法や高度な手法が統計解析または計算アルゴリズムで用いられた場合は、読者がそれらを再現できるように参考文献や使用例を添えて十分に説明すること。出版の際には計算に用いたコードやデータを Supplementary Information としてジャーナルが要求することもある。

■ いかなる結果でも盲目的に報告された旨、統計モデルが独立的に確認された旨については、下記の要領を踏まえた詳細な説明の添付が必要である。

1) 解析モデルやデータ処理のステップ、アウトプットの計算アルゴリズム、分類のために利用したカットポイントなどについての詳細。
2) それぞれの解析モデルや予測因子が完全に固定された日付。
3) 盲検化されたデータを保持し、評価を監督した人（たち）の指名。（例えば公平な仲介者など）
4) 解析モデルが固定された後に妥当性確認用データについて修正・追加・除外などが行われていない旨。また妥当性確認用データおよびその一部が、試験で利用されている解析モデルの評価・精錬には利用されていない旨。

④ The EMBO Journal

■ 利用した解析については、その解析についての仮定が満たされているかどうかを明確に記載しておく必要がある。特に正規性を仮定する統計手法を用いるときは、著者はどのように正規性の検証を行ったかを説明しなければならない。仮にデータがそれぞれの検定の仮定に従わない場合、ノンパラメトリックな手法を用いるべきである。

■ 症例数が小さい場合、正しい統計的な手法が用いられるべきであり、その妥当性も馳せて明記する必要がある。

⑤JAMA

■ 論文の Method において、十分に統計学的な知識のある読者であれば、解析に用いたデータセットを用いて再現できる程度の詳細を記載すること。特に、あまり一般的に知られていない手法を用いるときには、その手法が最初に紹介された正しい参考文献を添えて記載すること。より高度な手法や新たな手法が用いられる場合は、その手法と使い方についての簡単な説明を論文本文中に記載し、詳細をオンラインサプリメントに載せることを推奨する。

■ 要約統計量を計算するために利用した一般的に利用される解析手法については詳細を記載する必要はないが、Method の章で簡単に説明しておくべきである。

■ 無作為化研究・観察研究によらず主評価項目・副次的評価項目・サブグループや感度解析などの計画については計画段階で記載しておくことが求められる。計画時点で定められていなかった解析については「Post Hoc（後付け）」の解析であると明示する。

■ 無作為化比較試験において、すべての統計解析計画書については Method の章に記載し、オンライン Supplement で報告すること。また無作為化比較試験においては原則 Intention-to-treat の手法を用いて解析すること。完全な Intention-to-treat の法則から外れた場合は“modified intention-to-treat（改変された intention-to-treat）”である旨を記載し、どの点が改変されたかを明記すること。

■ Method の章の終わりには、統計的有意差を判定した基準などについて簡潔に記載すること。また解析に用いたソフトウェアについても、そのバージョン・製造者・利用したパッケージ名等を記載しておく。また解析手法としてソフトウェアのコマンド名などを直接記載しないこと（例: SAS proc mixed was used to fit a linear mixed-effects model）。解析に用いられたコードなどを記載する場合はオンライン Supplement を利用すること。

■ 解析はEQUATOR Reporting Guidelinesで記載された注意点を踏まえて、研究計画で記載された通り遂行し、後付けの解析はPost Hoc（後付け）であることを明示すること。