医療統計学リテラシー(4)

正しい統計検定手法の選び方

大阪市立大学大学院医学研究科
医療統計学教室
新谷　歩

しんたに　あゆみ

▶大阪市立大学大学院医学研究科医療管理医学講座医療統計学教授　▶PhD（米国 Yale 大学）▶国立病院機構本部総合研究センター生物統計室長，東海大学客員教授，名古屋第一赤十字統計アドバイザー，藤田医科大学統計アドバイザー，米国 Vanderbilt 大学客員教授，厚生労働省厚生科学審議会臨床研究部会委員・患者申出療養評価会議評価員　▶1996年 Yale 大学修士課程修了，2000年同大学博士課程修了，01年 Vanderbilt 大学医療統計学部講師，07年同准教授，14年大阪大学大学院医学系研究科臨床統計疫学寄附講座教授，16年より現職　▶1968年生まれ，兵庫県出身　▶著書：今日から使える医療統計，みんなの医療統計，あなたの臨床研究応援します　▶主な専門：ICU におけるせん妄研究，糖尿病，リウマチ，癌，感染症，腎臓病など多分野にわたる臨床データの統計解析。臨床研究ジャーナルに300篇以上に上る論文を執筆。最近では米国国立衛生研究所（NIH）推奨の電子データ集積システム REDCap の日本への普及に尽力　▶受賞：2013年 Vanderbilt 大学医学部ティーチングアワード

臨床研究，基礎研究を問わず，大切な研究結果をいざ世に送りだそうという時，統計解析で大変苦労した，査読者から解析が間違っていると指摘を受けて途方に暮れたという経験を持つ方は多いのではないだろうか。近年多くの国際ジャーナルでは統計専門家による査読が行われ，統計をいかに適切に行うかが研究の質を決めると言っても過言ではない。いかなるデータにどのような解析手法が用いられるかを知ることは，統計学を学ぶうえでの登竜門ともいえる。

図は米国国立衛生研究所（NIH）のある年の各疾病ごとの研究費を縦軸に，各疾病で失われた障害調整生命年を横軸に示している。ここで失われる生命年が大きい疾患ほど研究費が使われているというような相関関係が示せれば，米国 NIH は人命を救うために研究費を有効に支出していることになる。

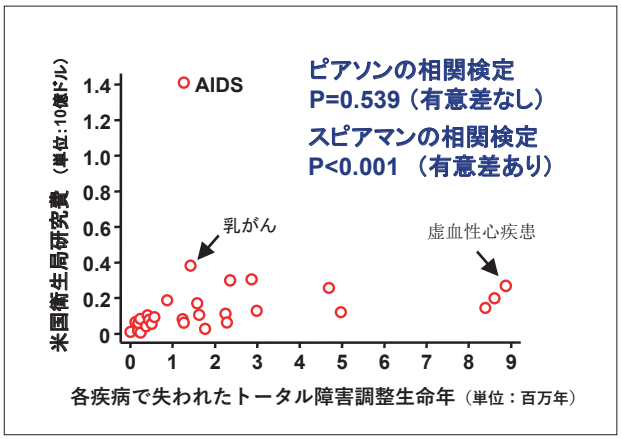

図　検定手法によって結果が変わる（参考文献１より作成）

ピアソンの相関検定では p 値は0.05を上回るので，相関なしという結果が出ている。一方，スピアマンの相関検定の手法を用いると p 値は0.0001より小さいという強い相関が検出された。ピアソンの相関検定は解析に用いられる変数が正規分布に従う必要があるが，研究費などのコストや年数のデータは，偏った分布をとるので，この場合は正規分布を仮定しないスピアマンの相関検定が正しい。この例のように，似たような検定でも使い方を間違えると，結果に大きく影響し，苦労して集めたデータも台無しになることがある。

著名な統計家かつ医師でもある Douglas Altman 氏は「誤った解析手法を故意にまたは知らず知らず使う，正しい解析手法を誤った方法で使用する，解析結果を間違って理解し発表するなどにより不当な結果を導く，このようなことが数え切れないほど多くの一般的，専門的な医学研究論文で日常的に行われていて，これは間違いなく医療スキャンダルだ」と言っている。**表**に正しい統計検定を選択する際の６つのポイントを挙げている。左から順に分類していけば最終的に正しい答えに到達する。大切なデータを正しく解析し，大切な研究結果を世に送りだせるよう，活用していただきたい。

それぞれの項目の分類の仕方をみてみよう。

（１）差を見るのか相関を見るのか？

治療群と未治療群で平均血圧を比べるといった具合に，差には複数の比較群が必要となる。一方，相関は複数の比較群は存在せず，年齢が上がれば血圧も上がると

表　統計手法を選択する際の６つのポイント（参考文献２より引用改変）

差 / 相関	比較データ間の対応性	変数の種類（正規性）	比較する群の数	サンプル数	適切な統計手法
差	対応なし	連続変数（正規分布）	2	総数 30 以上	スチューデントの t 検定
			＞2	1 群 15 以上	一元配図分散分析
		連続変数（非正規分布）/ 順序変数	2	制限なし	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 *
			＞2	制限なし	クラスカル・ウォリス検定 *
		２値変数	2	総数 20 未満	フィッシャーの正確確率検定 *
			≧2	総数 40 以上	ピアソンのカイ２乗検定※
		打ち切り例のある２値変数（生存時間解析）	≧2	イベント総数 10 以上	ログランク検定
	対応あり	連続変数（正規分布）	2	15 組以上	対応のある t 検定
			＞2	15 組以上	反復検定による分散分粧
		連続変数（非正規分布）/ 順序変数	2	制限なし	ウィルコクソンの符号彫位検定 *
			＞2	制限なし	フリードマン検定 *
		２値変数	2	制限なし	マクネマー検定
相関（関連性）		連続変数（正規分布）		総数 20 以上	ピアソンの相関係数
		連続変数（非正規分布）/ 順序変数		制限なし	スピアマンの順位相関係数 *
		２値変数		制限なし	ケンドールの順位相関係数 * カッパの相関係数（一致性）

＊ノンパラメトリック検定，それ以外はパラメトリック検定を示す。

いうような，２つの連続変数の関連を調べる。

（２）比較データが対応しているか？

新薬と既存薬に割り付けて効果を調べるような研究では，新薬を使う人と既存薬を使う人は全く無関係に選ばれるので比較群間のデータには対応がない。一方，ある目薬の効果を調べるため10人の患者のそれぞれ右目に新薬を左目は既存薬を投与するような研究では，右目と左目のデータは同じ患者のものであるため対応ありと見なされる

（３）アウトカム変数の種類は？

連続変数は年齢や血圧といった連続した値を持つ。性別（０：女性，１：男性）や生存（０：生存，１：死亡）のようなカテゴリーによって分類されたデータをカテゴリー変数と呼ぶ。カテゴリー変数の中で，性別のように２つの値しかとらないデータを２値変数とよび，３つ以上の値をとるカテゴリー変数は，順序付けができるかどうかで順序変数と名義変数に分類できる。重篤度（１：正常，２：中等度，３：重篤）などは順序変数であるが，病気の種類（１：癌，２：心臓病，３：感染症，４：糖尿病）は名義変数である。

（４）アウトカムが連続変数の場合その分布は正規分布であるか？

正規分布とは，データの分布が平均値に近い値の患者が一番多く，平均値から離れるに従って左右対称に数が減っていくような釣鐘型の分布のことをいう。大抵の医療データは正規分布には従わない。入院日数，入院費用，検査値，薬の使用量などはすべて歪んだ分布をとる。正規分布であればスチューデントの t 検定のようなパラメトリック検定，そうでなければマンホイットニー検定のようなノンパラメトリック検定が選ばれる。

（５）グループ間で比較を行うときグループの数は２つか３つ以上か？

血圧を表す変数が正規分布に従う場合，男女のような対応のない２群間で血圧を比べる場合はスチューデントの t 検定が使えるが，白人・黒人・アジア人など，対応のない３群以上の群間で比べる場合は分散分析（ANOVA）が用いられる。この場合，血圧の分布が正規分布に従わなければクルスカルワリス検定が用いられる。

（６）サンプルの総数は？

新薬と既存薬を投与した対応のない２群間で，生存者の割合（２値変数）を比べる場合は，総症例数が40以上であればピアソンのカイ二乗検定が使えるが，20未満の場合はフィッシャーの直接確率検定を使う（表※）。20以上40未満の場合は，新薬群の生存者と死亡者数，既存薬群の生存者と死亡者の数の１つでも５未満の数がある場合はフィッシャーの直接確率検定を使う。

最近では無料の統計ソフトの R をベースとした EZR（Easy R）などで，自動的に正しい検定手法を判断し，データの集計と比較を自動で行ってくれるありがたい機能もあり，以下の動画で詳細を見ることができる。

統計テストの選び方

データの集計比較表の自動作成の仕方

また統計ソフトを用いた解析には，データ入力の仕方を学ぶ必要があるので，以下のビデオも併せてご覧いただきたい。

データ入力の仕方

1）Gross CP, et al.：The relation between funding by the national institutes of health and the burden of disease. New Engl J Med, 340：1881 ～ 1887, 1999.
2）新谷 歩：今日から使える医療統計 . 医学書院 , 東京 , 2015.