

検定と標本数の関係 Test and Sample Size

高木廣文

看 護 研 究

第47巻 第2号 別刷

2014年4月15日発行

KANGO-KENKYU(The Japanese Journal of Nursing Research)

Vol.47 No.2/Mar. - Apr.2014

医学書院

第2回

検定と標本数の関係

Test and Sample Size

高木廣文 東邦大学看護学部長

はじめに

統計学をデータ解析に使用する目的は、その研究の目的に応じてさまざまかもしれませんが、ひとつだけ共通するところがあると考えられます。「糖尿病患者の新薬群と標準薬群を比較して、3か月後のグリコヘモグロビンの平均値に有意差があることを示したい」「新しい禁煙のための教育方法は従来の方法よりも、半年後の禁煙の継続率が有意に高率であることを示したい」、または、「セルフエフィカシーと職務満足度には有意な相関が認められることを示したい」などというように、統計学的に検定の結果が有意となる結論を得たいはずです。

研究の結果が統計学的に有意でない場合には、たとえ学術誌に投稿したとしても、何のエビデンスももたらさないのですから、通常はリジェクトされるのが普通でしょう。研究のために調査を行ない、収集したデータを解析し、その結果として、なんらかの研究上の価値ある結論を得るためには、少なくともデータの統計学的解析を通じて、主となる研究目的に関しては、有意な結果を得なければなりません。このためには、検定の特徴について、特に標本数と検定の関係を熟知しておく必要があるでしょう。

前回説明したように、統計学的仮説検定においては、「帰無仮説」と呼ばれる検定の目的に応じて、特別な仮説を設定します。その帰無仮説を否定することにより、2グループ間の母平均値に有意差があることや、2変数間に有意な相関があることを示したりします。

この帰無仮説を用いる検定では、2つのグループの母平均値に本当に全く差のない場合や、2つの変数の間に完全に相関のない場合以外には、必ず統計学的に有意となる結果を得ることができるのです。検定のもつ性質として、何となく納得しにくいかもしれません。しかし、この帰無仮説のもつ性質から導かれる結論は、研究デザインを考える上で、極めて重要な点になります。以下に、簡単に説明したいと思います。

標本数と p 値の関係について

表に、母相関係数の値が0.1, 0.2, 0.3の場合に、標本数がいくつあれば、p 値が有意になるかを、簡単な計算により求めた結果を示しました。母相関係数の検定は、通常は「無相関の検定」として知られており、t 検定を用いて行なわれます。これについては、本連載の第5回で詳細に解説します。ここでは、計算方法などは気にせずに、母相関係

数の大きさと標本数により、p 値がどのように変化するかを見てください。

実際には、母相関係数が 0.3 であっても、母集団から無作為に得られた標本から求めた相関係数が 0.3 にいつも一致するわけではありません。しかし、この例では標本でも母相関係数と等しい値が得られたものとして p 値を求めています。

表からわかるように、母相関係数が 0.3 の場合には、標本数が 40 人であれば p 値は 6% ですから、有意な相関とは言えないことになります。これが 2 倍の標本数である 80 人になると p 値は 0.7% ですから、1% 水準で有意な相関があると言えることになります。同様のことが、母相関係数が 0.2、0.1 の場合でも言えます。母相関係数が 0.2 の場合ならば、標本数が 160 人であれば p 値は 1.1%、母相関係数が 0.1 であっても標本数が 640 人であれば p 値は 1.1% ですので、やはり有意な相関として検定結果を示すことができます。ちなみに、表には示していませんが、母相関係数が 0.01 であっても、完全に無相関ではありませんので、標本数を 4 万人とすると p 値は 4.6% くらいになりますので、統計学的に有意な相関と認められます。普通の感覚から言えば、相関係数が 0.01 というのは、2 つの変数の間には相関がないと考えてよいような値です。しかし、無相関の検定では母相関が完全に 0 でなければ、検定結果が有意になることもあり得るのです。

何となく変な感じがするかもしれませんが、統計学的仮説検定法というのは、このような性質があるのだということは、肝に銘じておくべきです。このことから言えるように、検定で「有意」であることが、実際上の何か特別な「意味を有する」ということを示すものではないことも明らかだと思います。とはいえ、まずは有意な結果を得たければ、標本数は大きいに越したことはないということが言えるわけです。

表 母相関係数と標本数による p 値の変化

標本数	母相関係数		
	0.1	0.2	0.3
10	0.783	0.580	0.400
20	0.675	0.398	0.199
40	0.539	0.216	0.060
80	0.377	0.075	0.007
160	0.208	0.011	0.0001
320	0.074	0.0003	0.0000
640	0.011	0.0000	0.0000

検定における 2 つの誤り



統計学的仮説検定などと言うと、数値によって何もかもきっちりと正確に定められているように考えがちです。しかし、実際には意思決定のための方法ですので、その判断には誤りが必ずあり得ます。

検定では、帰無仮説が正しいと仮定した場合について、実際のデータが得られる確率を求めます。このときに、直接的に確率の計算をするのは大変難しいので、ある数値(「検定統計量」と呼ばれる)を計算し、その統計量が標準正規分布や t 分布などに従って分布すると仮定して、確率を求めています。通常は求めた確率が 5% より小さければ、そのような小さな確率のもとで実際のデータを偶然に得られることができるとは考えられないので、それはあり得ないことだと考えられます。あり得ないことが起きたとするのは、帰無仮説に基づく計算がおかしいのであり、そのような帰無仮説を棄却することになります。これにより、検定結果として統計学的有意性が言えることになるわけです。

しかし考えてみると、仮に検定結果として 4% の確率が得られたとすると、確かに確率は小さいのですが、100 回に 4 回はそのような現象が偶然起きてもよいことになります。したがって、この

確率は「帰無仮説が正しい場合に仮説を棄却する誤り」の大きさを示していることになります。

検定に関するこのような誤りは、「第1種の誤り(過誤)type I error」と呼ばれるものです。従来は記号としてはギリシャ文字の α (アルファ)でその確率を示すのが一般的でした。帰無仮説が正しいのに棄却する危険を表わす確率ですので、「危険率」などとも呼ばれることがあります。最近では、統計解析ソフトの出力で「p値」とか「p-value」とか呼ばれることも多くなってきています。表や文章中では、 $p = 0.04$ 、などとも表記されます。

検定のもう1つの誤りは、「帰無仮説が誤りなのに棄却できない」という誤りです。新しく開発された糖尿病薬は、本当に旧薬に比べて糖尿病のコントロールが上手くいくにもかかわらず、治療開始から半年後のグリコヘモロビンの平均値に、新薬群と旧薬群の2グループで有意差が認められなかったような場合が、そのような例になります。例えば、収集したデータに基づいて計算した確率が10%だった場合、10回に1回程度は偶然にそのような平均値の差が生じることになりますので、現在採用されているパラダイムからすると、5%より大きな確率では「有意差有り」という判断はできないことになります。

このような誤りは「第2種の誤り(過誤)type II error」と呼ばれるもので、研究上は絶対に避けたい誤りであると言えるでしょう。

第2種の誤りの確率は、通常はギリシャ文字 β (ベータ)で表わされます。検定の効率や研究デザインを考える場合、検定により有意な結果を得たいので、「帰無仮説が誤っている場合に棄却できる確率」、すなわち、 $1 - \beta$ 、の大きさが重要になります。これは「検出力 power of a test」と呼ばれます。実際の検定では、この検出力が大きいほど、研究で得たいと考えている有意な結果が確実に得られる可能性が高くなるわけです。したがって、研究を実施する前に、有意な検定結果を得やすくするために、検出力が高くなるように研究のデザインを設定する必要があります。

帰無仮説と対立仮説

AやBを母平均値や母相関係数とすると、帰無仮説は、基本的には「 $A = B$ 」という形式をしています。例えば、「2つのグループの肥満度は等しい」とか「2つの変数の母相関係数は0である」のような形式になります。

仮に、ある2変数間の相関について、帰無仮説「母相関係数 $\rho = 0$ 」が棄却された場合、われわれはどのように考えればよいのでしょうか。帰無仮説を否定した仮説、すなわち対立仮説として「母相関係数 $\neq 0$ 」が言えることは、検定の目的からしても明らかなことです。しかし、帰無仮説を否定した仮説としては、その他にも多くの対立仮説が考えられます。例えば、「母相関係数 $= 0.2$ 」や「母相関係数 $= 0.3$ 」などいろいろと考えることができます。

実は、対立仮説は研究上の作業仮説にもなり得ますので、研究開始前に考えることで、検出力の大きな検定を行なうための必要な標本数も決まってきます。

普通の検定では、帰無仮説 H_0 「 $A = B$ 」とすると、対立仮説はその否定ですから H_A 「 $A \neq B$ 」となります。このとき対立仮説 H_A は、AとBが平均値のような数値であれば、2つの対立仮説として、対立仮説 H_1 「 $A > B$ 」および対立仮説 H_2 「 $A < B$ 」のような形式に分けて考えることもできます。

ここで、2グループの母平均値に差があるかをt検定で解析する場合を考えてみましょう(具体的な計算式は、今後の連載で提示します)。2グループの母平均値を「母平均A」と「母平均B」とします。帰無仮説 H_0 は「母平均A = 母平均B」と書けます。これに対する対立仮説 H_A は「母平均A $\neq$ 母平均B」となります。2つの母平均が異なるというのは、一方が他方に比べて大きいということです。なので、この対立仮説 H_A は2つの対立仮説、 H_1

「母平均 $A >$ 母平均 B 」と H_2 「母平均 $A <$ 母平均 B 」の2つを含んだものだということがわかります。

両側検定と片側検定

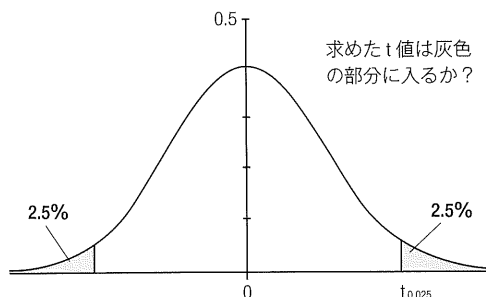
通常の検定では、どちらのグループの平均値や割合などが大きいかを問わずに検定を行いません。例えば、2つの母平均値の差の検定では、 t 値を計算する場合、標本から求めた2つのグループの平均値の差の計算で、 A から B を引く場合と、 B から A を引く場合の2通りがありますが、絶対値を計算することで、その順序を問題にしていません。

このように、帰無仮説 H_0 に対して対立仮説 H_A を考えて検定する場合、2つの対立仮説 H_1 「母平均 $A >$ 母平均 B 」と H_2 「母平均 $A <$ 母平均 B 」の2つを含んでいますので、検定の方法として「両側検定 both sided test, two tailed test」と呼ばれます。考え方としては、図1に示したように、検定に用いる分布の両裾で5%より小さな確率となるような統計量(例えば t 値)になるかで検定が行なえます。

対立仮説 H_A を2つに分けて考えることができるのならば、検定を行なうときに片側だけ考えて行なったほうが有意になりやすくなります。そのような考えで行なう検定は「片側検定 one sided test, one tailed test」と呼ばれます。

例えば、糖尿病の新薬の開発を行なったとき、新薬は標準薬よりコントロールが上手いくはずだから、 t 検定は片側でよいと考えるかもしれませんが、しかし、この考え方で片側検定するのは間違っています。新薬開発者や研究当事者はそのように考えても、疑い深い薬事審議会の委員や学術雑誌の査読者はそうは考えないでしょう。上手いくはずであるとか、新薬のほうが効果がある、というのは単なる思い込みであり、真実かどうか

図1 両側検定の考え方



は明らかではありません。

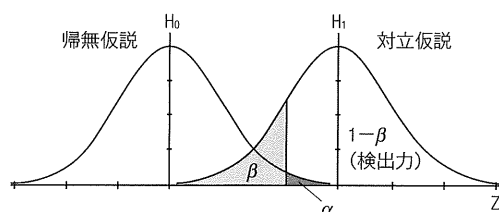
統計学的仮説検定は、あくまでも論理的に客観的に帰無仮説の成立・不成立を確認していくための方法論です。研究者の頭の中の考え方としては、片側検定はあり得ますが、実際に使われるのは両側検定です。基本的に片側検定は使わないと考えたほうがよいでしょう。

研究に必要な標本数とは

検定は帰無仮説を棄却するために行なうものです。したがって、帰無仮説は必ず棄却するように研究をデザインする必要があります。それでは、どのようにすれば検定の検出力が大きくなるのかというと、詳細な理論は省略するとして、主に2つの要因によって決まってきます。それはすでに表で見たように、①標本数、および、②効果の大きさ effect size, によってです。

検定と標本数に関しては、単純に言えば、標本数が大きければ大きいほど、検定結果は有意になりやすいと言えます。これは、帰無仮説の形式が「 $A = B$ 」のような形式をしており、それを否定することで、対立仮説「 $A \neq B$ 」を採択するという手順をとるためです。つまり、完全に A と B が母集団で一致していないのであれば、ほんのわずかでも違いがあるのであれば、そのわずかな差を見つけることは原理的に可能だからです。

図2 対立仮説と検出力



実際には、検定で用いる差というのは、平均値や相関係数などから求めた統計量になりますが、その統計量の標準誤差との比が問題になります。

例えば、2つのグループの母平均値の差の有無を検定したいのであれば、母平均値の差 Δ とその標準誤差 σ の比 (Δ/σ) が重要です。この比は「効果の大きさ(効果サイズ)」と呼ばれ、この値が大きいほど標本数は少なくともよく、小さいほど標本数が多く必要になります。実際に必要な標本数は、第1種の誤り α と第2種の誤り β にも依存します。

図2に対立仮説と検出力の関係を示しました。実際には、対立仮説は無数に存在しますので、どのような対立仮説を設定するかで検定の検出力は大きく異なります。さらに、効果サイズをどのくらいに設定するかで、検定に必要とされる標本数が定まってきます。

実際の計算では、第1種の誤り α を5%とし、第2種の誤り β を α の4倍の20%、すなわち検

出力は80%に設定して、研究に必要な標本数を決定します。詳細については、今後、それぞれの検定の説明のときに、解説を加えていきたいと思います。

おわりに

今回は検定にまつわる話題として、帰無仮説と対立仮説、検定に関する誤りと検出力、そして研究に必要な標本数に関して、簡単な解説を加えました。研究倫理審査委員会などでも、研究に必要な標本数をどのように決めたのかは、問題になることがあります。このような問題解消のために、次回以降は具体的な検定の方法を解説し、そのような検定を行なう場合に必要とされる標本数をどのように定めるかを説明したいと考えています。

●文献

- 高木廣文、林邦彦(2006). エビデンスのための看護研究の読み方・進め方. 中山書店, pp.45-51.
- 高木廣文(2009). ナースのための統計学, 第2版. 医学書院, pp.50-51.
- 高木廣文(2014). 統計学のキー・ポイント「検定」に焦点を当てて第1回検定の基本的な考え方. 看護研究, 47(1), 52-58.

たかぎひろふみ●東邦大学看護学部
〒143-0015 東京都大田区大森西 4-16-20

NURSING BOOK INFORMATION

医学書院

医療者のためのExcel入門

超・基礎から医療データ分析まで

田久浩志

●B5 頁200 2013年
定価: 本体2,300円+税
[ISBN978-4-260-01845-6]

Excelの基本的なキー操作といった“超・基礎”から、便利な集計機能「ピボットテーブル」を駆使したデータの集計・分析、グラフ作成のポイントまで、豊富な画面例でやさしく伝授。医療現場を想定したサンプルデータを用いているので、身近な業務データの整理にも即役立つ！「スライド原稿を作る便利ワザ」など、医療者がより便利にExcelを使いこなすための有益な情報も満載。