連載 EBNのための統計の読み方⑬

サンプルサイズの定め方

高木廣文・林　邦彦

EBNURSING Vol.4 No.1 2004〔別刷〕　中山書店

EBNのための統計の読み方⑬

サンプルサイズの定め方

高木廣文 TAKAGI Hirofumi
新潟大学医学部保健学科

林　邦彦 HAYASHI Kunihiko
群馬大学医学部保健学科

はじめに

統計学に関する質問のなかでも，頻繁に出てくるものの一つは，「この研究にはどのくらいの標本数が必要ですか」というものではないだろうか．このような質問の背景には，統計学的仮説検定は，標本数によって有意になったりならなかったりするという事実を，多くの研究者が知るようになったということからだろう．たとえば，ある研究で新しい治療法と従来の治療法の効果に，統計学的に有意差が認められなかったからといって，それが2つの治療法の効果に完全に差がないことを示すものではないのである．後述するように，その研究での標本数より2倍の標本数で検定すれば有意になるかもしれないし，それでもだめなら，さらに倍の標本数で研究すれば，統計学的に有意な結果を得られるかもしれないのである．

すでに説明したように，統計学的仮説検定の結果は，有意水準と検出力の影響を大きく受けており，標本数の不足や偶然の変動により，期待した結果が得られないことがある．この点から，適切で必要最小限の標本を確保するように研究計画を立てる必要がある．しかし，実際には必ずしも研究のすべてが，このような仮説検証的な研究ばかりではない．多くの研究は，実態調査的であったり，仮説探索的なものであることのほうが普通ではないだろうか．そのような研究でのサンプルサイズは，どのように決めればよいのだろうか．

今回は，検定と標本数の関係をまず簡単に解説し，さらに，研究に必要なサンプルサイズについて，いくつかの場合に分けて考えてみたい．

統計学的仮説検定と標本数の関係

まず，検定結果が標本数の影響をどの程度受けるものかを理解するために，簡単な例を示そう．

❶は，ケース・コントロール研究で，喫煙の胃癌に対する影響を調べたものである（架空例）．胃癌患者20例中の喫煙者は10例（50％），一方，コントロール群では20例中5例（25％）であった．簡単な計算から，オッズ比$=\frac{10\times15}{10\times5}=3$，となり，喫煙による胃癌のリスクは3倍と推定できる．しかし，偶然の変動によってこのような結果が得られたのではないこ

	喫煙		計
	あり	なし	
患者群	10人	10人	20人
対照群	5人	15人	20人
計	15人	25人	40人

❶喫煙と胃癌の関係

とを示すためには，統計学的仮説検定を行う必要がある．

検定の帰無仮説は「母集団での胃癌発生に喫煙は無関係である」というものであり，これは「母集団でのオッズ比（母オッズ比）＝1」を検定することである．計算は，通常の四分表のカイ2乗検定を用いればよい．イェーツの修正カイ2乗値χ_Y^2を求めると，

$$\chi_Y^2=\frac{40\times(|10\times15-10\times5|-\frac{40}{2})^2}{20\times20\times15\times25}=1.707$$

となる．この値は，自由度1のカイ2乗分布に従うので，有意確率p（p値）を求めると，$p=0.191$となる．p値は5％よりも大きいので，この結果は有意ではない．すなわち，帰無仮説を棄却することができないので，喫煙の胃癌に対する影響は，統計学的にあるとはいえないことになる．よくある解釈上の誤りに，帰無仮説を棄却できないことを，帰無仮説が正しいものと考えて，「喫煙は胃癌に影響しないことが証明された」とするものがある．しかし，これは完全に間違いなので注意が必要である．

❷は，❶の標本数をそのまま2倍にしたものである．この場合も，オッズ比は3のままである．しかし，イェーツの修正カイ2乗値χ_Y^2を求めると，$\chi_Y^2=4.320$，となり，$p=0.0377$，となる．すなわち，p値は5％よりも小さく，帰無仮説は棄却される．したがって，帰無仮説を否定し，「喫煙は胃癌に影響がある」という結論を得ることになる．

このように，比較する母集団間での差や変数間の関係が完全にない場合（差が0，相関係数が0など）以外は，標本数さえ大きくすれば，統計学的検定で必ず有意な結果を得ることができるのである．一般に，まったく差がないとか，相関が完全に0であるような健康事象はまれなので，検定結果が有意でないというのは，単に標本数が少ないためであるという可能性が高いことになる．ただし，統計学的な有意性が必ずしも現実的に意味のあることを示すわけではない．したがって，むやみに標本数を増やそうと努力するのではなく，その研究に適切な標本数を収集するように研究計画を作成する必要があるだろう．

	喫煙		計
	あり	なし	
患者群	20人	20人	40人
対照群	10人	30人	40人
計	30人	50人	80人

❷標本数が2倍の場合

研究目的とサンプルサイズ

研究に必要な標本数は，検証したい仮説に依存している．したがって，研究の目的が仮説検証的な場合には，その仮説に応じた適切な標本数を求めなければならない．一方，研究が仮説や理論探索的な場合，検定は2次的なものになるだろう．この点から，研究に必要なサンプルサイズは，研究目的が，(1) 仮説検証，(2) 仮説探索，かによって分けて考えなければならない．

◆仮説検証的研究での標本数の定め方

前号で説明したように，仮説検定には2種類の誤りがある．すなわち，“第1種の誤りα”と“第2種の誤りβ”である．通常，第1種の誤りは「有意水準」とよばれており，第2種の誤りは，$1-\beta$を考えて，検定の「検出力」として知られている．

検出力は，帰無仮説が正しくない場合に，帰無仮説を棄却する確率の大きさを示すものである．したがって，仮説検証的研究では，帰無仮説が誤っている場合に必ず棄却できるようにするためには，検出力を十分に大きくするように標本数を決めねばならない．しかし，一般の検定においては，対立仮説を明確に定義できないため，第2種の誤りβの大きさを

定めにくい．そのため，検出力も求めにくい場合が多い．Cohenによると，βはαの約4倍程度が望ましいものとされている[1]．それに従うと，有意水準が5％であれば，第2種の誤りは20％（検出力は80％），有意水準が1％であれば，第2種の誤りは4％（検出力96％）と設定することとなる．

実際のサンプルサイズを求める式は，やや複雑であり，ここでは簡単な例のみにとどめるが，

(a) 有意水準αの大きさ

(b) 検出力$1-\beta$の大きさ

(c) 対立仮説での効果の大きさなど

を事前に決める必要がある．上記のうち，有意水準や検出力は，5％と80％などと，研究者によって任意に定めることができるが，問題は(c)である．

仮説検定では，帰無仮説は「母相関係数＝0」「2群の母平均値は等しい」「喫煙の胃癌発生のオッズ比は1」などのように，相関の大きさや差の大きさなどを定めないで，まったく存在しないものとして検定する．しかし，検出力の計算には，対立仮説として「母相関係数＝0.2」「2群の母平均値の差＝10」「喫煙の胃癌発生のオッズ比は3」などのように，実際に数値を推定しなければならない．さらに，使用する変数の母分散の情報が必要な場合もある．これらの数値がわからない場合には，研究に必要な標本数を求めることは不可能である．

比較研究に必要な標本数を求めるための近似的な計算式として，以下のような簡単だがすぐれた式が知られている[2]．

2グループA，Bの母平均値の差の検定のためには，まず2つの母平均値の差をΔとする．2グループの母分散を等しいものと仮定し，それをσ^2とする．このとき，検定の有意水準を両側2α（片側α），検出力を$1-\beta$とすると，各グループの標本数nは，

$$(1)\quad n=2(z_\alpha+z_\beta)^2\cdot\frac{\sigma^2}{\Delta^2}+\frac{z_\alpha^2}{4}$$

によって近似的に求めることができる．ここで，z_αとz_βは，それぞれ標準正規分布の上側α点と上側β点の値である．

なお，この式は，2群の特定事象の割合の差（母比率の差）の検定にもそのまま応用できる．すなわち，2グループの割合の差Δ（$=p_1-p_2$）を定めることができれば，母割合をpとすると，$\sigma^2=p(1-p)$，$p=\frac{(p_1+p_2)}{2}$と置くことで前出の(1)式を用いることができる（p_1とp_2は，2つのグループでのある特性の割合）．

計算に必要とされる，$\lambda=\frac{\Delta}{\sigma}$は検定すべき差が標準偏差の何倍であるかを示す値であり，「エフェクト・サイズ ；effect size」とよばれることがある．この値が大きいほど標本数は少なくてよく，逆に小さな場合には多くの標本数が必要になる．

通常は，有意水準は片側2.5％（両側5％），βをその4倍とすると片側検定での検出力は90％となる．このとき，計算に必要な標準正規分布での各数値は，$z_{0.025}=1.96$，$z_{0.1}=1.282$となる．$z_{0.025}$はほぼ2なので，前述の(1)式はより簡単にすると，

$$(2)\quad n=\frac{21.02\cdot\sigma^2}{\Delta^2}+1$$

となる．したがって，近似的な標本数としては，

$$(3)\quad n=\frac{21\cdot\sigma^2}{\Delta^2}$$

で十分ということになるだろう．

高血圧群と健常者群での食塩摂取量が，それぞれ平均値17g/日，14g/日であるとすると，この3g/日の差を検出するためには，標準偏差の推定値が必要になる．食塩摂取量の標準偏差を5g/日と推定すると，各グループで必要な標本数nは(2)式を使うと，

$$(4)\quad n=21.02\times(\frac{5}{3})^2+1=59.4$$

となり，各群60例の標本が必要とされる．(3)式の場合には58.3となるが，切り上げて同様に約60例とすればよいだろう．

前述の胃癌と喫煙の例では，患者群での喫煙割合

は50％（$p_1=0.5$），健常群では25％（$p_2=0.25$）であった．有意水準が片側2.5％，検出力90％の場合，必要とされる各群の標本数は，まず，

母割合$p=\frac{(0.5+0.25)}{2}=0.375$として，

$$\lambda^2=\frac{(p_1-p_2)^2}{p(1-p)}=\frac{(0.5-0.25)^2}{0.375\times0.625}=0.2667$$

である．したがって，

$$n=\frac{21.02}{\lambda^2}+1=\frac{21.02}{0.2667}+1=79.8$$

となるので，各群80例の標本数が必要となる．

前述の例では，各40例で検定結果は有意となっているのだが，この計算式からは2倍の標本数が必要という結果になった．この理由は，通常の検定では検出力は50％に設定されているということによる．すなわち，$z_{0.5}=0$なので，この値を（1）式に代入して，必要な標本数を計算すると，

$$n=\frac{2\times(1.96+0)^2}{0.2667}+1=29.8$$

となり，各群30例でよいことになる．ただし，実際に観察される人数に端数はないので，仮に健常者群の喫煙者が7人観察された場合には，補正なしのカイ2乗値は4.593となり，$p=0.0321$で5％で有意となる．しかし，1人増えて8人観察された場合には，カイ2乗値は3.455となり，$p=0.063$で，有意とはならない．なお，イェーツの修正カイ2乗値では，どちらの場合も有意にはならない．

現実の調査研究では，偶然性の変動の影響があるので，検出力を高く設定しないと思ったような結果を得られないことがあり得る．したがって，検出力を50％としている単純な検定の式が必要とする標本数よりも，より多くの標本数が必要となる点に注意しなければならない．

◆探索的研究で必要とされる標本数

サンプルサイズの問題は，仮説検定の要求によって生じる問題であることは，これまでの説明で理解できたものと思う．したがって，検証すべき数量的な仮説がない研究では，一般に標本数を適切に設定する便利な公式は存在しない．それでは，まったく適当に標本数を決めてよいかというとそうでもなく，ある程度は考慮すべき点がいくつかある．まず，

標本数は多いほどよい

ということである．すなわち，30例の研究よりも300例の研究のほうがよいということである．大部分の統計学的推定では，推定値の信頼性は，ほぼ標本数の平方根に比例している．したがって，母平均値や母割合の信頼区間を1/10の幅にするためには，10倍ではなく約100倍の標本が必要になる．

現実には，それほど多くの標本調査は不可能なことも多いが，質問紙調査などでは，

調査項目数の2倍程度の標本数

はほしいところである．標本数は多いほどよいのだが，逆に少なすぎると一般的な回答パターンが得られている保証がなくなってしまう．

特に，分析に多変量解析を使用する場合には，理論的に解が得られなくなってしまうことがある．すなわち，

因子分析や主成分分析を用いる場合，少なくとも分析に用いる変数の個数以上の標本数を確保する必要がある

できれば，変数の個数の2倍以上の標本数が最低でも必要だろう．さらに，

逆行列の計算を含むような解析方法，たとえば重回帰分析などを用いる場合には，絶対に変数の個数以上の標本数が必要である

上記のように，探索的研究では，仮説検定のための事前情報がないので，標本数に関しては確定的な方法はない．しかし，多くの看護研究では，唯一の数量的な仮説検証を目的とするような場合はめったにないのではなかろうか．そのような場合には，少なくとも調査項目数の2倍程度の標本数を集めるようにするべきだろう．

おわりに

今回は，研究で必要とされるサンプルサイズについて簡単に説明した．誌面の都合や複雑な計算式を避けたために，多くの検定方法に関しての詳細な説明はしていない．より詳しい説明に興味がある読者は，成書や下記の参考文献などを参照していただきたい．

文献

1) Cohen J : Statistical Power Analysis for the Behavioral Sciences. Academic Press ; 1969.
2) 竹内啓：確率分布と統計解析．日本規格協会 ; 1975. p.107.

参考文献

Fleiss JL 著，佐久間昭訳：計数データの統計学．東京大学出版会 ; 1975.
林邦彦，高木廣文：EBN のための研究計画―統計学的推論：推定と検定．EB NURSING 2003 ; 3(4) : 82-87.
高木廣文，林邦彦：コントロールや無作為化はなぜ必要なのか．EB NURSING 2002 ; 2(4) : 102-107.
椿広計，藤田利治，佐藤俊哉編：これからの臨床試験．朝倉書店 ; 1999. p.171-172.
上坂浩之：臨床試験における被験者数設計の視点と方法．計量生物学 2003 ; 24(1) : 17-41.
柳井晴夫，高木廣文：臨床試験における標本数の定め方．医学のあゆみ 1977 ; 102(13) : 881-886.
柳井晴夫，高木廣文編著：新版看護学全書 統計学．メヂカルフレンド社 ; 1995. p.88-105.