
原 著

医学における統計学の応用について

— (第5編) 検出力 $1-\beta$ 及び必要標本数について —

新潟大学医学部衛生学教室 (主任: 山本正治教授)

遠藤 和男・山崎 理

An Application of Statistical Methods to Medical
Science (Part V) on Power ($1-\beta$) and Sample Size

Kazuo ENDOH and Osamu YAMAZAKI

Department of Hygiene and Preventive Medicine

Niigata University School of Medicine

(Director: Prof. Masaharu YAMAMOTO)

We Japanese are generally too sensitive to the statistical significance level (α)¹⁾ and very credulous to believe that 5 percent should be an absolute standard under any condition. Although there is no room for discussion on the importance of type I error; α , type II error; β or power; $1-\beta$ should be also taken into consideration²⁾.

Recently, the importance of statistical power has been pointed out even in the field of medical science. We are therefore requested to refer to power²⁾³⁾ or the sample size, when we especially conduct a clinical trial⁴⁾⁻⁶⁾.

Here, the authors put an emphasis on 2 points as follows :

- 1) Methods for computing power; $1-\beta$ exactly after testing by significance level; α .
- 2) How to estimate the sample size satisfying the significance level and power, before starting a clinical trial or an animal experiment.

Key words: significance level (α), power ($1-\beta$), type I and II error, sample size

有意水準 α , 検出力 $1-\beta$, 第 I 種及び第 II 種の過誤, 標本数

Reprint requests to: Kazuo ENDOH,
Department of Hygiene and Preventive
Medicine, Niigata University School of
Medicine, Niigata City, 951, JAPAN.

別刷請求先: 〒951 新潟市旭町通1番町
新潟大学医学部衛生学教室 遠藤 和男

はじめに

特に日本人研究者は、有意水準 $\alpha^1)$ には敏感であり、どのような場合にも5%が絶対な基準であると信じ込んでいるかも知れない。第I種の過誤 α の重要性については議論の余地はないが、これまで第II種の過誤 β または検出力 (power) $1-\beta$ については、ほとんど考慮されてこなかった²⁾。

ただし最近、医学の分野においても検出力の重要性が見直されてきた。したがって、医学論文を投稿する際にも、検出力²⁾³⁾ や特に臨床試験における必要標本数⁴⁾⁶⁾ についての記載が求められる場合も増えてきた。

そこで本編では、以下の2点に力点を置いて解説を加える。若手研究者の参考になれば幸いである。

- 1) 有意水準 α で検出力 $1-\beta$ を正確に求める方法。
- 2) 臨床試験や実験を始める前に、有意水準及び検出力を満足する必要標本数をいかにして求めるか。

I. 第I種及び第II種の過誤について

1. 第I種の過誤 (type I error) と危険率 α

通常、研究者が気にしている危険率 p は、第I種の過誤¹⁾ を犯す確率 α 、すなわち、帰無仮説 H_0 が真の場合に、これを誤って棄却してしまう確率である。

初心者が統計学的な検定方法になじめないのは、統計学が直接的な証明ではないためである。例えば、A群とB群との平均値の差を検定する際、本当は $\bar{x}_A > \bar{x}_B$ (または $\bar{x}_A < \bar{x}_B$) を証明したいのに、わざわざ逆説的に帰無仮説 $H_0: \bar{x}_A = \bar{x}_B$ を設定し、これを棄却することで、間接的に対立仮説 $H_1: \bar{x}_A \neq \bar{x}_B$ を採用する。間接的証明なので、棄却の際に誤りを犯す確率を p として示す。本当に証明したい H_1 が真の時、これを直接的に検定する方法は、今のところ知られていない。

また、帰無仮説が真である場合、これを正しく採用する確率は $1-\alpha$ である。上下の棄却域を $\alpha/2$ と設定すると、両側 α はいわばハズレの確率、信頼度 $1-\alpha$ はアタリの確率に相当し、裏腹の関係になっている。データが正規分布に従う時、統計量 $\pm z(\alpha/2)$ はそれぞれ、

信頼区間の上・下限を示している (図1)。

なお、有意水準 α を常に5%に設定してよいかどうかについては、拙著¹⁾ を参照されたい。

2. 第II種の過誤 (type II error) と危険率 β

第I種と逆の過誤を犯す確率である。すなわち、帰無仮説 H_0 が偽である場合に、これを誤って採用してしまう確率である。一般的には危険率 β より検出力 (power) : $1-\beta$ の方が有名であるが、正しく理解されているかどうか、はなはだ疑問である。

そこで、上記の帰無仮説 $H_0: \bar{x}_A = \bar{x}_B$ について、以下の①~③の条件のもとに検討してみる。

- ① データが正規分布に従い、等分散とみなしてスチューデント¹⁾ の t 検定を適応したら、 H_0 を棄却できた。
- ② 正規性の有無を検討せず、マン-ホイットニーのU検定¹⁾ を用いると、 H_0 を棄却できなかった。
- ③ 当初 H_0 の真偽は不明であったが、後に H_0 が偽であることが判明した。

②では偽の H_0 を誤って採用したことになり、第II種の過誤 β に該当する (表1)。逆に①では偽の H_0 を正しく棄却した。つまり、真の対立仮説 H_1 を正しく採用したことになる。一般的に正規性が仮定できる場合、ノン・パラメトリック技法の方が検出力が低い¹⁾⁷⁾ ので不利であると言われている。

- ちなみに、
- ④ データが正規分布に従わない場合で、
 - ⑤ H_0 が真の場合について検討すると、

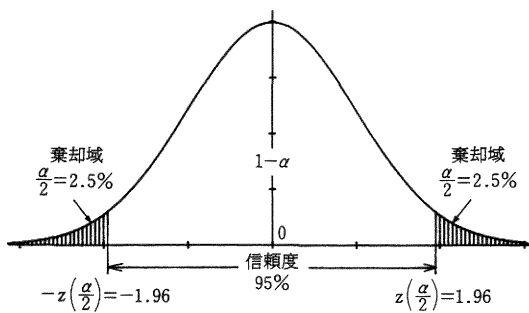


図1 両側の棄却域 α と信頼度 $1-\alpha$ との関係

表1 第I種と第II種の過誤の関係

判 断		帰無仮説 H_0 (: $m = \bar{x}$ など)	
帰無仮説 H_0	対立仮説 H_1	真 [対立仮説 H_1 は偽]	偽 [対立仮説 H_1 は真]
採用する	棄却する	正しい判断 ($1-\alpha$)	第II種の過誤 (危険率 β)
棄却する	採用する	第I種の過誤 (危険率 α)	正しい判断 (検出力 $1-\beta$)

今度は①で H_0 を棄却した判断は誤りとなる。このように、有意差の出やすい統計方法を恣意的に選択する過ちを、第Ⅲ種の過誤⁸⁾と呼ぶことがある。

3. 第Ⅰ種と第Ⅱ種の過誤の関係

第Ⅰ種の過誤は、棄却してはならない真の H_0 を早トチリして過剰に捨ててしまった、いわばアワテ者の誤りであり、第Ⅱ種の過誤は、本来棄却すべき偽の H_0 をボンヤリと見逃して残しておいた、いわばウスノロの誤りと解説⁹⁾されている。

両者の関係は図示した方がわかりやすい。すなわち α が↓なら β は↑(図2)、 α が↑なら β は↓(図3)となる。この関係は、スクリーニング検査での感度と特異度との関係¹⁾ に似ていて、残念ながら両者を同時に小さくすることはできない。

ただし、平均値の差が同じであっても、2群の標本数が多くなれば、確率 p は小さくなる事実を知っている研

究者も多いであろう。 α を固定した場合、標本数を多くすればある程度 β を小さく、つまり検出力 $1-\beta$ を大きくすることができる。有意水準 α を両側に設定した場合、 β は α の2~5倍でよいとされる¹⁰⁾¹¹⁾ が、絶対的な基準があるわけではない。

また、 $\alpha=5\%$ で H_0 が棄却され、 H_1 が採択されたことを裏返すと、真である H_1 (H_0 は偽) を95%以上の確率で正しく採用したことになるため、検出力は不要であると勘違いする人がある。通常は H_0 の真偽が不明であるという、大前提を忘れているための議論にすぎない。

Ⅱ. 検出力 $1-\beta$ の算出方法

1. 母平均と標本平均との比較

A. 母分散が既知の場合

実例は多くないが、全国規模の調査等で母平均 m 、母

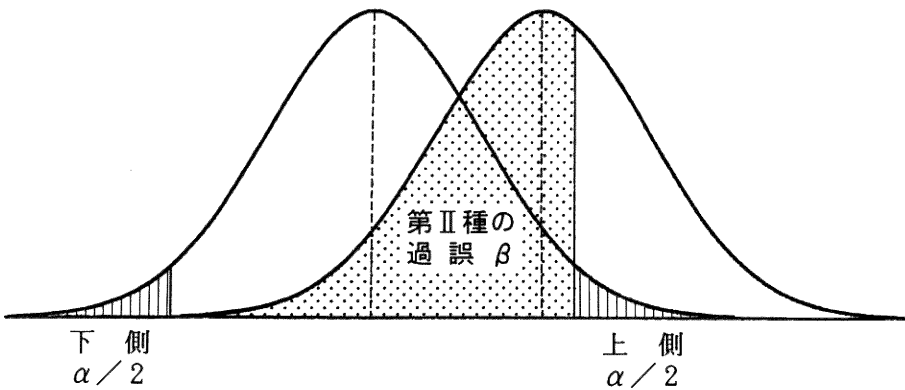


図2 第Ⅰ種の過誤 α (||||) ↓, 第Ⅱ種の過誤 β (●●) ↑の場合

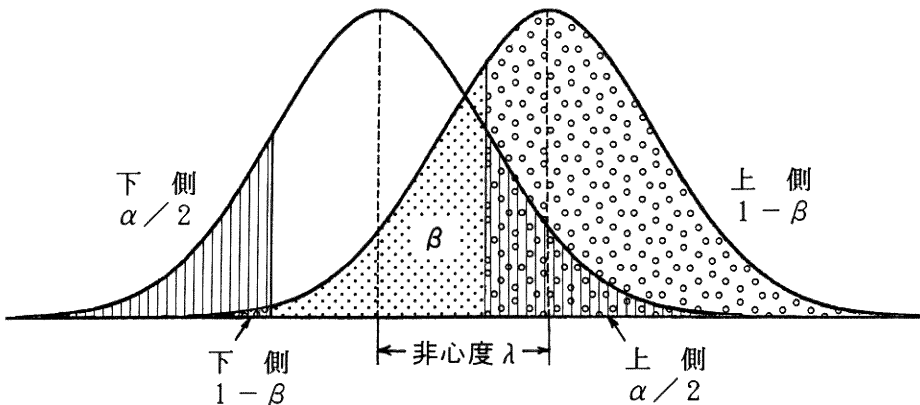


図3 α (||||) ↑及び β (●●) ↓の場合, 両側 $1-\beta$ との関係

分散 S^2 が知られている場合である. n 個の標本について

標本平均を \bar{x} とすると, $\begin{cases} \text{帰無仮説 } H_0: m = \bar{x} \\ \text{対立仮説 } H_1: m \neq \bar{x} \end{cases}$
 検定式は帰無仮説のもとに, 標準正規分布に従う.

$$z_0 = \frac{m - \bar{x}}{S/\sqrt{n}} \sim N(0, 1^2) \dots\dots\dots \text{式 I)}$$

z_0 の絶対値 $> z(\alpha/2)$ なら, 有意水準 α で H_0 を棄却して, 有意差ありと判断 ($p < 0.05$ と表記) する.

ところで対立仮説のもとでは,

$$\lambda = \frac{m - \bar{x}}{S/\sqrt{n}} \sim N(\lambda, 1^2) \dots\dots\dots \text{式 II)}$$

非心度 λ だけ標準から右にズレた分布に従う (図 3).

標準正規分布において, 統計値 z で与えられる下側の累積確率を $P(z)$, 上側の累積確率を $Q(z)$ とした場合, 検出力 $1 - \beta$ は両側の確率で, α と λ の関数である¹²⁾.

$$1 - \beta = P[-z(\alpha/2) - \lambda] + Q[z(\alpha/2) - \lambda] \dots\dots\dots \text{式 III)}$$

通常 $1 - \beta$ は上記の値に 100 を乗じて % で示される. ところで $\alpha = 5\%$, $z(\alpha/2) = 1.95996$, $\lambda = 2$ の場合, 式 III) の第 1 項は $P[-3.95996] = 0.0037628\%$ となる. すなわち, $-z(\alpha/2) - \lambda < -4$ の場合, 第 1 項はほとんど無視しても差し支えないであろう. また, 正規分布の確率は下側から求めるプログラムが多いため, 第 2 項は以下のように書換えられる.

$$1 - \beta \approx P[\lambda - z(\alpha/2)] \dots\dots\dots \text{式 IV)}$$

【例 1】(拙著¹⁾ p 46~を参照のこと.)

一般健康診査の成績によると, 60~69 歳男性について最大血圧の平均値 \pm 標準値 (mmHg) は,

$\begin{cases} \text{X 県} & 135.87 \pm 19.12 \\ \text{A 町 (121 人)} & 139.5 \pm 21.2 \end{cases}$ であった.

有意水準 = 5% として検出力を求めなさい.

$$\lambda = \frac{|139.5 - 135.87|}{19.12/\sqrt{121}} = 2.088 > z\left(\frac{0.05}{2}\right) = 1.96$$

したがって, A 町の平均値は X 県より有意 ($p < 0.05$) に高い. 式 III) を用いて検出力を計算する.

$$1 - \beta = 0.00259411\% + 55.0926\% = 55.0952\%$$

したがって, 検出力 $1 - \beta = 55.1\%$ となる.

B. 母分散が未知の場合

母分散 S^2 は未知であるから, 標本から求めた不偏分散 u^2 で代用するほかない.

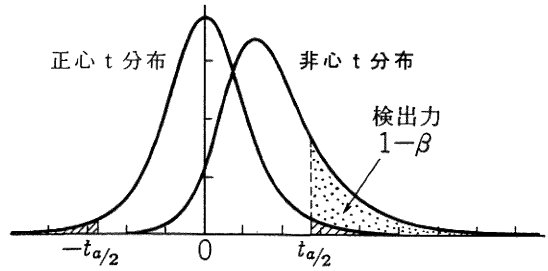


図 4 正心 t 分布と非心 t 分布との関係

$$\lambda = \frac{m - \bar{x}}{u/\sqrt{n}} \dots\dots\dots \text{t 検定式 V)}$$

[自由度 $df = n - 1$]

λ の値を式 III) または IV) に代入して $1 - \beta$ を算出するのは適当ではない. t 分布が左右対称であるのに対して, $1 - \beta$ は非心度 λ だけ t 分布が右にズレただけではなく, 図 4 に示したとおり右にスズを引く分布に従う.

これを非心 t 分布 (noncentral t distribution) と呼ぶ.

この累積確率を正確に求めるのは至難のワザであるが, 幸いにして左右対称の正心 t 分布の確率を用いる, 以下のような近似式¹³⁾ が知られている.

$$x = \frac{\lambda - t(\alpha/2)d}{\sqrt{1 + t(\alpha/2)^2(1 - d^2)}} \dots\dots\dots \text{式 VI)}$$

ただし, $\begin{cases} t(\alpha/2) \text{ は自由度 } df \text{ の正心 } t \text{ 分布の値} \\ d = 1 - 1/4df + 1/32df^2 \end{cases}$

$1 - \beta = P[x]$ として下側の累積確率を求める. 詳細は略すが, λ を有効数値 4 桁で与えてやると, $1 - \beta$ が 50% 以上の時には, 小数点以下 1 桁まで信頼できる. 50% 未満及び自由度が小さい場合は多少近似が悪くなってしまいが, 誤差はせいぜい 0.3% くらいである.

【例 2】例 1 で X 県の平均値だけが知られているのに, 標準偏差 19.12 が不明の場合を考えてみる. 標本の標準偏差 $u = 21.2$ を用いることになる. $\alpha = 5\%$ として,

$$\lambda = \frac{|139.5 - 135.87|}{21.2/\sqrt{121}} = 1.883 < t_{120}\left(\frac{0.05}{2}\right) = 1.980$$

したがって, 両者の間に有意差は認められない. また, まず d の値から計算して, 式 IV) を用いると,

$$d = 1 - 1/480 + 1/32 \cdot (120)^2 = 0.9979$$

$$x = \frac{1.883 - 1.98 \cdot 0.9979}{\sqrt{1 + 1.98^2(1 - 0.9979^2)}} = -0.0921$$

$$1-\beta=P[-0.0921]=46.331\% \approx 46.3\%$$

このように、A. 母分散が既知の場合に誤って式V)を用いると、本来認められるべき有意差を検出できなくなる恐れがあるが、例2の方が検出力が小さいためであると理解できよう。

2. 2つの標本平均の比較

等分散のF検定後、スチューデント(Student)のt検定¹⁾が実行できる場合、式V)の非心度λを以下のよりに求めた上で、式VI)を用いればよい。

$$\lambda = \frac{\bar{x} - \bar{y}}{u / \sqrt{1/n_x + 1/n_y}} \quad [df = n_x + n_y - 2] \quad \dots\dots\dots \text{式VII)}$$

uは共通の分散¹⁾(pooled variance)の平方根

2つの標本平均の比較は実際に非常に多く用いられているため、様々な条件下で適用できる検出力曲線が示されている²⁾¹²⁾¹⁴⁾。しかし、曲線から正確な1-βを求めるのは不可能であり、単なる目安にすぎない。

なお、等分散に用いるF分布にも非心F分布¹²⁾¹³⁾があり、F検定の検出力も算出できるが省略する。また、ウェルチ(Welch)のt検定¹⁾についても、計算されたt値を非心度λとおき、自由度φを計算して近似式VI)を適用すればよい¹²⁾。

【例3】最大血圧値について以下のようなデータが得られた。α=5%として検定しなさい(拙著¹⁾p68)。

	人数	平均値	標準偏差
A群	16	150.0	31.34
B群	32	132.0	26.11

F検定の結果、等分散とみなしうる。

等分散とみなせるので、共通の分散u²及び式VII)からλを求める。計算は略すが、u²=779.71(u=27.92)

$$\lambda = 2.106 > t_{46} \left(\frac{0.05}{2} \right) = 2.013 \quad \left\{ \begin{array}{l} d = 0.9943, \\ x = 0.1024, \end{array} \right.$$

したがって、A群の方が有意に(p<0.05)平均値が高いと言える。また、1-β=P[0.1024]≈54.0%となる。

しかし、F検定を実施せずにウェルチ法を用いると、

$$\lambda = 1.979 < t_{26} \left(\frac{0.05}{2} \right) = 2.056 \quad \left\{ \begin{array}{l} d = 0.9904, \\ u = -0.0551 \end{array} \right.$$

1-β≈47.8%となり、スチューデント法よりも検出力が落ちる。つまり、等分散のF検定を実行せずにウェ

ルチ法を用いるのは、誤りであると理解できよう。

3. 母比率と標本比率の比較

全国調査などが実施されている場合は、母比率Pが既知であると考えてよい。標本数がnで陽性者数がxの場合、標本比率はx/nであるが、イエーツ(Yates)の修正項を含んだ正規近似式¹⁾が計算されていれば、式III)を用いて容易に1-βを算出できる¹²⁾。

$$\lambda = \frac{|x - nP| - 0.5}{\sqrt{nP(1-P)}} \quad [nP \geq 5 \text{ の場合}] \quad \dots\dots\dots \text{式VIII)}$$

期待値nP<5の場合、正確にはF検定式またはポアソン分布¹⁾¹²⁾によるが、イエーツの修正が実施されている場合、誤差はそれほど大きくない。

【例4】平成6年度小学校の齲蝕罹患率は88.01%であった。ある小学校1~6年生142人を調べたところ、虫歯、治療歯及び喪失歯のまったくない児童は30人であった。α=5%として検討しなさい。

実際の罹患率は142-30=112、期待値=142×0.8801=124.97>5から正規近似式を適用する。

$$\lambda = \frac{|112 - 124.97| - 0.5}{\sqrt{124.97(1 - 0.8801)}} = 3.223 > z \left(\frac{0.01}{2} \right)$$

したがって、この学校の罹患率は有意に(p<0.01)に低い。また -λ-z(α/2)<-4から式IV)を用いて、

$$1-\beta = P[3.223 - 1.96] = 89.6623\%$$

4. 二群の標本比率の比較

A. カイ二乗検定値を用いる方法

カイ二乗検定において、√χ²(α)=z(α/2)の関係を思い出せば、計算は容易である。実際にはイエーツの修正項を含んだχ²値¹⁾を用いるので、この平方根を非心度λと考えると、式III)またはIV)を用いる¹²⁾。

【例5】▷拙著¹⁾p86~を参照のこと。

A, B2薬の薬効試験について検討する(α=5%)。

効果	あり	なし	計	有効率
A薬	8	14	22	36.4%
B薬	2	22	24	8.3%
計	10	36	46	21.7%

$$\chi^2_c = \frac{(|8 \cdot 22 - 14 \cdot 2| - 46/2)^2 \cdot 46}{22 \cdot 24 \cdot 10 \cdot 36}$$

$$= 3.78130 < \chi^2_1(0.05) = 3.841$$

したがって有意差なしと判断される(p=0.0518287)。

$\sqrt{\chi^2}=1.94456$ であるから式Ⅲ)を用いて、 $1-\beta=49.3901\%$ と計算される。

B. フィッシャー (Fisher) の直接確率法

2×2 表の4つの期待値のうち、少なくとも1つが5未満の場合に適用される。通常両側の確率に直すため、求めた確率を2倍して評価するが、2倍する前の上側確率を与える標準正規分布のz値をλと考えて、式Ⅲ)またはⅣ)を用いばよい¹²⁾。

【例6】例5の期待値を求めると、

8の期待値=22・10/46=4.78<5 となるため、本来の例5はフィッシャー法の適応である。

詳細は略すが、 $\Sigma p(i)=0.024738$ であるから、有意性の判定としては $2 \times \Sigma p(i)=0.049476 < 0.05$

上側確率0.024738を与えるz値=1.96447をλとして式Ⅲ)に代入すると、 $1-\beta=50.1846\%$ となる。

C. その他の計算方法

① 標本での出現比率の平均値 $\bar{p} = \frac{2 \text{ 群の陽性数の計}}{2 \text{ 群の標本数の計}}$

から

$$z = \frac{|p_1 - p_2|}{\sqrt{\bar{p}(1-\bar{p})(1/n_1 + 1/n_2)}} \dots\dots\dots \text{式IX)}$$

をλとして式Ⅲ)を用いる方法¹¹⁾が紹介されているが、この式はイエーツの修正がされておらず、 $\sqrt{\chi^2}$ を用いた場合と同じ結果を得る。検出力だけでなく、検定にあっても誤差が含まれる¹⁾。

例5では、 $1-\beta=64.0580\%$ となり、誤差が大きい。

② アークサイン変換による方法

比率の分散分析の場合、分散が標本数から解放されるので用いられる方法⁹⁾¹²⁾であるが、特に2群の標本数が大きく異なる場合には、誤差が大きくなってしまう。

例5では、 $1-\beta=67.0655\%$ となる。

以上、Aの検定結果は若干Bとは異なるものの、p値の誤差もわずかであり、検出力の値もC①及び②の方法より、フィッシャー法の値に近いので推奨できる。

5. その他の検出力について

A. 相関係数の検出力

① 標本数nにかかわらず $|r| \geq 0.75$ の場合は、フィッシャーのZ変換値をλとして、式Ⅲ)またはⅣ)を用いる。

② $|r| < 0.75$ で標本数 $n < 120$ (拙著¹⁾では<50の場合)の場合は、t値をλとして式Ⅵ)を用いる。

③ $|r| < 0.75$ で標本数 $n \geq 120$ (拙著¹⁾では ≥ 50

の場合)の場合は、z値をλとして式Ⅲ)またはⅣ)を用いる。

B. 分散分析法の検出力

等分散のF検定と同様に非心F分布を用いるが、難し過ぎるので、専門書¹²⁾¹⁵⁾を参照されたい。

C. ノン・パラメトリック検定技法の検出力

ノン・パラメトリック検定技法でも標本数 ≥ 20 くらいなら、期待的平均値及び分散を求めて、式Ⅲ)から検出力を計算できる。パラメトリック技法に比べて検出力は劣るが、詳しくは専門書¹²⁾¹⁶⁾を参照されたい。

Ⅲ. 必要標本数nの算出方法

以下主として、第I種の過誤 $\alpha=5\%$ 、第II種の過誤 $\beta=20\%$ 、すなわち検出力 $1-\beta=80\%$ の場合を例として述べていくことにする。なお、両側 α を与える正規分布のz値を $z(\alpha/2)$ [$\alpha=5\%$ の時、1.95996] とおく。式Ⅳ)では $1-\beta$ は上側確率なので、下側の確率になおして $z(\beta)$ [$\beta=20\%$ の時、0.841261] が問題となる。

1. 母平均と標本平均との比較

A. 母分散が既知の場合

式I), II)ではいずれも $\frac{|m-\bar{x}|}{S}$ が、標本数nに

影響を与えていることがわかる。 $\delta = \frac{|m-\bar{x}|}{S}$ とおく

と、 δ は標本平均 \bar{x} が母平均mと比較して、母標準偏差の何倍離れているかを示している。必要標本数をnとすると、

$$n \geq \left[\frac{z(\alpha/2) + z(\beta)}{\delta} \right]^2 \dots\dots\dots \text{式X)}$$

nは α 、 β 及び δ の関数である¹²⁾ことがわかる。

【例7】例1の検出力は55.1%であった。80%以上とするためには標本数はいくつ必要か？

$$\delta = |139.5 - 135.87| / 19.12 = 0.189854$$

式X)から、 $n \geq 217.70$ が計算できる。

整数をとり、必要標本数は218である。式Ⅲ)を用いて検算してみると、 $1-\beta=80.0441\%$ ($p=0.005060$)となる。

母平均及び母分散は既知なので、これから求める標本平均 \bar{x} が、どのくらいになるか予測することが重要である。予測のつかない場合、 $\delta=0.1\sim 0.5$ の範囲でシミュレーションすることが必要である。

B. 母分散が未知の場合

検出力の計算では、非心t分布を用いたが、必要標本数の計算では、基本的に式X)を用いたあとで、次表に示した補正值 Δ1 を用いて補正すればよい(表2)¹²⁾。

【例8】例2の検出力は46.3%であった。母分散は未知として80%以上となる必要標本数を求めなさい。

$$\delta = 0.171226 \text{ を式X)に代入して, } n \geq 267.64$$

補正值 Δ1=2 であるから、必要標本数は270である。検算すると、 $1-\beta=80.053\%$ ($p=0.005096$) となる。

しかし、両群の標本平均及び不偏分散は実際にデータを集めてからでないとわからないので、過去の文献のデータによるか、またはシミュレーションするしかない。

2. 2つの標本平均の比較

式VII)において標本数が同じで $n_x=n_y=n$ の場合、

$$\lambda = \frac{\bar{x} - \bar{y}}{u} \sqrt{n/2} \quad [df=2(n-1)]$$

となるから、式X)を2倍すればよいことになる。

また、2群の標本数が異なる場合には、調和平均¹⁾を

とると $\frac{1}{n} = \frac{1}{n_x} + \frac{1}{n_y}$, 変形して $n = \frac{2n_x n_y}{n_x + n_y}$ なので、数の

多い群が少ない群のk倍の人数である場合、

$$c = 2k / (1+k) \quad \dots\dots\dots \text{式XI)}$$

を係数としてかければよいので、式X)から、

$$n \geq c \left(\frac{z(\alpha/2) + z(\beta)}{\delta} \right)^2 \quad \dots\dots\dots \text{式XII)} \quad [\text{同数なら } c=2]$$

として、最終的に表2の補正值 Δ2 を加える¹²⁾。

【例9】例3は多い群は少ない群の2倍であった。

$$\delta = |150.0 - 132.0| / 27.92 = 0.644699$$

$$c = 2 \cdot 2 / 3 = 1.33333 \text{ から } n \geq 25.1786$$

表1の補正值 Δ2=1 であるから、少ない方は27人、多い方は54人とすればよいことになる。しかしながら、式III)を用いて検算すると、 $1-\beta=77.0338\%$ にすぎない。29人と58人(あわせて87人)の場合に79.9674%とほぼ満足する値が得られる($p=0.005765$)。

念のため2群が同数の場合は39人(あわせて78人)でよく、さらにA, B群とも標準偏差 $u=27.92$ ならば、 $1-\beta=80.2556\%$ ($p=0.005674$) である。すなわち、スチューデントのt検定では、両群の例数及び標準偏差が等しい時に、検出力が最も高い¹²⁾ことが理解できるであろう。実験計画を立てる際に例数を揃えるのは、以上に述べた理由による。

表2 必要標本数の算出の際の補正值

両側 α	z 値	z ² /2	補正值 Δ1	補正值 Δ2
0.001	3.291	5.415	6	3
0.005	2.576	3.318	4	2
0.01	1.960	1.921	2	1
0.05	1.645	1.353	2	1
0.10	1.282	0.822	1	1

3. 母比率と標本比率の比較

ほとんどの教科書はこの点について触れていない。式VIII)でイエーツの修正項がない場合には、標本平均 $p_s = x/n$ を用いれば下記のように変形できる¹⁷⁾。

$$\lambda = \frac{|p_s - P| \sqrt{n}}{\sqrt{P(1-P)}} \text{ から, } \delta = \frac{|p_s - P|}{\sqrt{P(1-P)}}$$

を式X)に代入すればよい。ただし、イエーツの修正項を含んでいないため、以下の補正值¹⁸⁾を加える。

$$\Delta 3 = 1 / |p_s - P| \quad \dots\dots\dots \text{式XIII)}$$

【例10】例4の必要標本数を求めてみる。

標本比率 $p_s = (142-30)/142 = 0.7887$ であるから、

$$\delta = \frac{|0.7887 - 0.8801|}{\sqrt{0.8801(1-0.8801)}} = 0.2814$$

これを式X)に代入して補正值 Δ3 を加えると、

$n \geq 99.83 + 1/0.0914 = 110.77$ から111人でよい。罹患患者数は $111 \times 78.87\% = 87.5457$ であるから、式VIII)及びIII)から検出力を検算できる。

{ 罹患患者が87人なら、 $1-\beta=84.8604\%$ ($p=0.002904$)
 罹患患者が88人なら、 $1-\beta=76.5950\%$ ($p=0.007241$)
 このように比率の場合、陽性者数は整数であるためピタリと80%になることは少ない。必要標本数はやや多めにとった方が無難であろう。

4. 2群の標本比率の比較

2群の標本平均の比較について、両群の標本数が等しい場合に検出力が大きくなることを示した。2群の標本比率 p_1, p_2 の場合も同様で、両群の標本数が等しい場合には検出力が最大になる¹²⁾だけでなく、式IX)の平均的な比率 $\bar{p} = (p_1 + p_2)/2$ が容易に計算できる。

$$\text{したがって, } \delta = \frac{|p_1 - p_2|}{\sqrt{P(1-P)}}, c=2 \text{ として式XII)に代}$$

入すればよい。ただし、イエーツの修正項を含んでいないため、最終的に以下の補正值¹⁸⁾を加える。

$$\Delta 4 = 2 / |p_1 - p_2| \dots\dots\dots \text{式 X IV)}$$

$\Delta 3$ の係数が 1 なのに、 $\Delta 4$ の係数が 2 であるのは、式 XI) で $c=2$ であることと同じ理由による。

【例 11】ある疾患の有病率は過去の統計から 7% と予想できたが、A 地区では最近多いという情報を得た。対照の B 地区とともに何人ずつ調べたらよいか？ A 地区 = 12%，B 地区 = 7% として必要標本数を求めなさい。

標本数が同じなら、 $\bar{p} = (7+12) = 9.5\%$ である。

$$|0.12 - 0.07| = 0.05, \delta = 0.1705,$$

式 XI) 及び X IV) から、 $n \geq 539.99$ 、 $\Delta 4 = 40$ なので、一群の標本数を 580 人にすればよいことになる。

期待される有病数は A 地区 $580 \times 0.12 = 70$ 、B 地区 = 41 であるが、幸いにして差はちょうど 5% と変わらない。 $\sqrt{z^2 c} = 2.795$ を式 III) に代入して、 $1 - \beta = 79.8152\%$ ($p = 0.005194$) とほぼ満足する結果を得る。

IV. 検出力が重視される理由

例 1, 3, 6 ではいずれも $p < 0.05$ で帰無仮説 H_0 を棄却して対立仮説 H_1 を採用したが、検出力はせいぜい 50% を超えたにすぎない。なぜなら 5% の場合には、対立仮説 H_1 が真の場合に、これを積極的に採用したとは言えないからである。つまり、検定は「間接的な証明に過ぎない」ことが理解できるであろう。

5% の有意差が認められたという論文に対して、有意差が認められなかったとする追試論文²⁾をよく見かける。例 1, 3, 6 の場合、追試結果の 40% 以上は「有意差なし」として報告されることが予想される。追試と結果が異なるようでは、普遍的な真理とは言いがたい³⁾。

したがって最近、臨床試験を実施する際に $1 - \beta$ を考慮に入れて、必要標本数をシミュレーションすることが多くなってきたのである⁵⁾¹⁹⁾。 α だけしか記載していないと、計画段階で $1 - \beta$ を考慮しなかったとして、あとから非難を受ける場合も多くなった。

例 7～例 9 では $p < 0.01$ なので、初めから有意水準 α を 1% に設定すればよいのでは？ という疑問も生じるであろう。しかし式 X) では n は α 、 β 及び δ の関数になっているので、 α が変われば当然 n も変化する。

【例 12】例 7 で $\alpha = 1\%$ の場合の必要標本数を求める。 $z(0.01/2) = 2.57583$ を式 X) に代入する。

$n = 325$ と例 7 よりかなり多い。したがって $p < 0.01$ なら検出力も 80% に近いと考えてよいが、始めから $\alpha = 1\%$ と設定する必要はない。

また、ギリギリ $p > 0.05$ となった場合の検出力は低いので、症例数が少ないためなのか？ 本当に差がないの

か？ 検出力の検討が必要とされる²⁾。帰無仮説 H_0 を採用して有意差なしと判定しても、積極的に意味がないとは言えないとして、レフィリーからクレームがくることもある⁴⁾。仮に例数を倍として計算しなおすとよい。有意差が認められた場合には、単に例数が少ないために、本来検出されるべき差が認められなかったと判明する。

参 考 文 献

(訳本については [] 内に原著を示した)。

- 1) 遠藤和男, 山本正治: 医統計テキスト, 西村書店(新潟), 1992.
- 2) 斧田大公望, 訳: 医薬統計学入門, メディカルリサーチセンター(東京), pp 103~126, 1996. [Glantz, S.A.: Primer of biostatistics, 3rd ed., McGraw-Hill (New York), 1992.]
- 3) 橋 敏明: 医学・教育学・心理学にみられる統計的検定の誤用と弊害, 医療図書出版(東京), 1986.
- 4) 津谷喜一郎, 折笠秀樹, 監訳: 医統計学の活用, pp 238~251, サイエンティスト(東京), 1995. [Bailer, J.C. and Mosteller, F., ed.: Medical uses of statistics, 1st ed. Massachusetts Med. Soc. (Mass.), 1986.]
- 5) コントローラー委員会, 監訳: クリニカルトライアルよりよい臨床試験を志す人たちへ, pp 174~179, 183~185, 篠原出版(東京), 1989. [Pocock, S.J.: Clinical trials: A practical approach, John Wiley & Sons (New York), 1986.]
- 6) 重松逸造, 柳川 洋, 監修: 新しい疫学, pp 125~128, 日本公衆衛生協会(東京), 1991.
- 7) 市原清志: バイオサイエンスの統計学—正しく活用するための実践理論, 南江堂(東京), 1990.
- 8) 千葉大学統計グループ, 訳: ケンドール統計学用語辞典, 丸善(東京), 1987. [Kendall, M.S. and Buckland, W.R.: A dictionary of statistical terms, 4th ed., Longman Group (London), 1982.]
- 9) 上村 桂: 医統計学, 文永堂(東京), 1979.
- 10) 竹内 啓, 編: 統計学辞典, 東洋経済新報(東京), 1989.
- 11) 古川俊之, 監修, 丹後俊郎: 新版医学への統計学, pp 182~199, 朝倉書店(東京), 1993.
- 12) 統計数値表編集委員会, 編: 統計数値表, JAS-1972, 表 pp 323~371, 解説 pp 163~184, 日本規格協会(東京), 1972.

- 13) 統計数値表編集委員会, 編: 簡約統計数値表, pp 55~62, 146~151, 日本規格協会(東京), 1985.
- 14) 佐久間昭(日本消化器病学会, 編): 医学統計Q&A, pp 17~22, 1987.
- 15) 田中 豊, 垂水共之, 編: パソコン統計解析ハンドブック, III 実験計画法編, 共立出版(東京), 1986.
- 16) 白旗慎吾, 編: パソコン統計解析ハンドブック, IV ノンパラメトリック編, 共立出版(東京), 1987.
- 17) 柳川 洋, 編: 疫学マニュアル, 第5版, pp 124~125, 南山堂(東京), 1996.
- 18) **Fleiss, J.L.:** Statistical methods for rates and proportions, 2nd ed., John Wiley & Sons (New York), 1981.
- 19) 重松逸造, 監訳: 疫学・臨床医学のための患者対照研究—計画の立案・実施・解析, ソフトサイエンス(東京), 1985.
[**Schlesselman, J.J.:** Case-control studies; design, conduct, analysis, Oxford Univ. Press (London), 1982.]

(平成8年10月3日受付)
