

連載 統計学のキー・ポイント—「検定」に焦点を当てて・6 [最終回]

対応のある場合の2グループの有意差検定
Tests for the Difference Between Matched Pair Groups

高木廣文

看 護 研 究

第47巻 第7号 別刷

2014年12月15日発行

KANGO-KENKYU (The Japanese Journal of Nursing Research)

Vol. 47 No. 7/Nov. - Dec. 2014

医学書院

第6回

対応のある場合の2グループの有意差検定

Tests for the Difference Between Matched Pair Groups

高木廣文 東邦大学看護学部長

対応のある場合とは？

高血圧症の患者グループに減塩指導を行なったとします。実施した減塩指導は本当に効果があったのかを、どのように調べればよいでしょうか。1つの方法として、指導を受けた患者グループについて、指導前と指導後の塩分摂取量を測定し、指導後の塩分摂取量が指導前よりも少なくなっていれば、実施した指導は有効であったと考えてよいでしょう。または、最終的な目的は血圧を下げることですから、指導の前後で血圧を測定し、実際に血圧が下がっていれば、減塩指導は有効であったと考えてよいでしょう。

同様に、ダイエットの効果を知りたいければ、体重を測定し、以前よりも減っていればダイエットは成功したものと考えてよいでしょう。また、学生に講義を行ない、その講義が有効であったかは、講義内容に関連する知識量の増加が1つの目安になります。講義の効果を調べるには、同一の内容の試験を講義の前後に行ない、講義の前後での得点の増加を調べることで、その講義の評価が行なえるはずです。

このように、ある対象に対して特定の介入を行なった場合、その介入の効果は、関連すると考えられる項目を繰り返し測定し、そのデータの差を

表1 データの分類と使用する検定

データの分類	使用するデータ	検定
量的	前後の測定値	対応のある場合のt検定
	前後の符号	符号検定
質的	前後の符号と順位	Wilcoxonの符号付順位和検定

比較することにより、評価することができます。そのようなデータは、各対象について同一項目が繰り返し測定されるため、通常は「対応のあるデータ」と呼ばれます。実際には、対応のあるデータでは時間経過による介入以外の他の要因の影響もあり得るので、適当なコントロールを設定するほうが、より適切な検討が行なえます。この点については後述しますが、まずは単純な対応のある前後の比較を説明したいと思います。

対応のあるデータとしては、上記のような塩分などの摂取量や、体重のような量的な場合や、ある疾患の臨床症状の有無や軽減などのように質的な場合もあります。対応のあるデータを扱う方法は、表1に示したように、データが量的か質的かによって、主に3つの方法がよく使われています。それらについて、以下に順に説明することになります。

対応のある量的データの 場合



表2は僧帽弁狭窄症患者8名について、PTMC（経皮的僧帽弁交連切開術）を行なう前後の僧帽弁口面積（ cm^2 ）を心エコーにより求めたデータです（架空データ）。このデータから、PTMCは僧帽弁口面積の拡張に有効といえるでしょうか。

各ケースについて同一の調査や測定を繰り返して行なう場合、最もよく用いられる方法は、「対応のある場合の母平均値の差の検定（ t 検定）」と呼ばれる方法です。

まず以下に簡単に、理論的な解説をします。

いま、ケース数を N とし、減塩指導などの介入（処理）の前後での変数 X の測定値を、それぞれペアのデータとして、 (x_1, y_1) , (x_2, y_2) , ..., (x_N, y_N) と書くことにします。なお、処理や介入としては「ある薬を服用させる」「手術をする」「食事指導を行なう」「禁煙教育をする」などさまざまなケースが考えられます。これらの処理の効果は、その介入前後での変数の測定値に差があるか否かにより、その効果の有無を判断することができます。当然ですが、どのような変数について測定するかは、その変数が注目している処理に関係している必要があります。

介入前後の測定値の差を、

$$d_i = x_i - y_i, \quad (i=1, 2, \dots, N)$$

としたとき、差の平均値 \bar{d} 、および差の不偏分散 s^2 をまず求めます。

検定の帰無仮説は、「母集団での差の母平均値 δ は0」です。差の平均値 \bar{d} の標準誤差は、 s/\sqrt{N} により求められます。差の平均値の標準化をすると、 $\delta=0$ なので、検定のための統計量は、

$$t_0 = \frac{\bar{d} - \delta}{(s/\sqrt{N})} = \frac{|\bar{d}| \times \sqrt{N}}{s}$$

と求められます。

表2 PTMC 前後の僧房弁口面積

No	前	後	差
1	1.0	1.5	0.5
2	1.1	2.0	0.9
3	0.6	1.1	0.5
4	0.5	1.0	0.5
5	0.7	1.1	0.4
6	1.2	1.6	0.4
7	1.1	1.2	0.1
8	0.8	0.9	0.1

この統計量 t_0 は、自由度 $N-1$ の t 分布に従います。統計量 t_0 と自由度から、検定のための有意確率（ p 値）を計算し、 p 値が 0.05（5%）か 0.01（1%）以下ならば、帰無仮説を棄却し「有意差あり」とし、 p 値が 5% 以上ならば「有意差なし」となります。

有意差が認められた場合には、行なった処理に効果があったものと考えてよいことになります。なお、この検定では、分析を行なう変数の分布が、治療や指導などの介入により、値の大きい方向（または小さい方向）に、全体として一定の「ずれ」が生じたか否かを検定しているので、処理の前後の測定値間には相関が認められるはずですが、もしも前後の測定値間に相関が認められない場合、この検定方法は適切ではないことになります。

【例題 1】

表2のデータについて、上記の t 検定を行なってみます。

表2の最も右側の欄に示したように、まず2回の測定値の差を求めます。次に、差の平均 \bar{d} と差の不偏分散 s^2 を求めます。

これらの値は、

$$\bar{d} = (0.5+0.9+\dots+0.1)/8 = 0.424$$

および、

$$s^2 = [(0.5 - 0.424)^2 + \dots + (0.1 - 0.424)^2] / 7$$

$$= 0.065$$

と求められます。

検定のための統計量 t_0 は、

$$t_0 = \frac{|\bar{d}| \times \sqrt{N}}{s} = \frac{0.424 \times \sqrt{8}}{\sqrt{0.065}} = 4.714$$

となります。この t_0 は t 分布に従い、その自由度は、 $8 - 1 = 7$ ですので、実際に p 値を計算すると、 $p = 0.0022$ と極めて小さいことがわかります。

したがって、これらの結果から PTMC の前後での僧帽弁口面積には有意差が認められ、PTMC は有効なものと考えてよいことになります。なお、PTMC 前後の僧帽弁口面積の単相関係数を求めると、0.725 となり、かなり高い相関を示しています。したがって、本方法による検定は妥当なものと考えてよいでしょう。

ここでは、特定の処理の前後での測定値の比較という立場で説明を行ないましたが、疫学研究などで用いる研究デザインのうち、患者-対照研究（または、曝露-対照研究）などでも、この方法は有効です。すなわち、ある疾患の患者と性・年齢・職業などの条件を一致させたケースとコントロールをペアとして設定し（マッチング）、マッチさせなかった各種の特性についての比較を行なうことも可能です。

符号検定 sign test による比較

測定されたデータの数値自体の信頼性が低い場合や、ある症状の有無などのように、もともとデータが質的な情報の場合に、いわゆる「分布によらない検定（ノンパラメトリック検定）」により、特定の処理や介入の効果を検定することができます。母集団での変数の分布に正規分布などの特定の分布を仮定しないノンパラメトリック検定は、データ

の数値にそれほど信頼がおけない場合でも、データの大小関係程度の情報を利用して検定が効率よくできるため、広く用いられています。

いま、ペアを組んでいる一方のグループを A、もう一方を B とします。2つのグループでの A、B の差を調べるために用いることができる情報は、特定の変数についての量的な測定値であってもよいし、各ペアについて、「A のほうが B より有効である」「A より B の症状がよくなった」といった質的なものでも構いません。各ペアについて、A より B がよい結果を与えている場合に「+」とし、逆の場合を「-」とします。どちらも同一の結果（0 の「タイのある場合」と呼ぶ）の場合には、検定のための情報がないので以後の分析から除外します。

正（+）を示したペアの数を N_+ とし、負（-）を示したペアの数を N_- とします。符号検定では、正と負のペアの個数のみを用いるので、検定の対象となる全ペア数を新たに N とすると、

$$N = N_+ + N_-$$

となります。

1. 符号検定による正確な確率

検定のための帰無仮説は、「処理の効果はない」、すなわち「+か-は確率 1/2」という帰無仮説になります。処理の効果がなければ、各ペアで結果が正になるか負になるかはどちらの可能性も等しく、確率は 1/2 となるはずですが、したがって、 N 個中 N_+ 個以上が正となる確率を p_0 とすると、これは N 個中 N_- 個以下が負となる確率と一致し、次のような 2 項分布により求められます。すなわち、

$$p_0 = \sum_{k=N_+}^N \frac{N!}{(N-k)! \times k!} \times \left(\frac{1}{2}\right)^N$$

$$= \sum_{k=0}^{N_-} \frac{N!}{(N-k)! \times k!} \times \left(\frac{1}{2}\right)^N$$

により求められます。

上式から、測定値間の差の計算の方向や、+- の符号の付け方は、結果に無関係であることがわ

かるでしょう。ただし、ここで求められた確率 p_0 は、一定方向への偏りについて計算されていますので、普通は p_0 を 2 倍した値を、検定の有意確率として用います(両側検定)。

符号検定は、このように測定値の大小、もしくは結果の良否といった質的な情報のみを用いて検定を行なっているため、数値をそのまま用いる t 検定に比べ、一般に検定の効率はよくありません。もとのデータが本質的に質的なものでない場合には、特別な理由がない限り、後述するウィルコクソン(Wilcoxon)の符号付順位和検定か、前述の t 検定を用いるほうがよいでしょう。

2. 正規近似による p 値とマクネマー検定

符号検定により、有意確率を正確に求めるための計算は、標本数 N が大きくなるとやや煩雑になってきます。そのような場合、標準正規分布による近似により、検定を行なうと便利です。

いま、 $N_+ > N_-$ とします。検定のためには、

$$z_0 = \frac{2N_+ - N - 1}{\sqrt{N}} = \frac{N_+ - N_- - 1}{\sqrt{N_+ + N_-}}$$

を求めます。

z_0 は、 N すなわち標本数が大きければ、標準正規分布に従うようになります。ここで、上式の z_0 の計算では、正規分布への近似がよりよくなるように「連続性の補正」を行なっています。検定の方法は、 z_0 の値が 1.96 より大きければ、両側 5% 水準で有意差ありとし、同様に 2.58 より大きければ、1% で有意差ありとすればよいのですが、正規分布近似により p 値を求めて、 $p < 0.05$ ならば「有意差あり」とすればよいでしょう。

なお、 z_0 を 2 乗した統計量は自由度 1 のカイ 2 乗分布に従い、それに基づいて行なう検定は「マクネマー(McNemar)検定」と呼ばれています。すなわち、

$$\chi_0^2 = \frac{(N_+ - N_- - 1)^2}{N_+ + N_-}$$

を求めます。この χ_0^2 は自由度 1 のカイ 2 乗分布に従いますので、 χ_0^2 が 3.84 より大きければ 5% 水準で、6.63 より大きければ 1% 水準で、それぞれ「有意差あり」とすればよいです。または、近似的に p 値を求めて、 $p < 0.05$ ならば「有意差あり」とすればよいでしょう。

【例題 2】

表 2 のデータを符号検定し、PMTC の有効性を検討してみましょう。

表 2 では、8 ペアのデータの差はすべて + なので、計算は 8 個中 8 個すべてが正になる確率を求めればよいでしょう。

$$\begin{aligned} p_0 &= \sum_{k=8}^8 \frac{8!}{(8-k)! \times k!} \times \left(\frac{1}{2}\right)^8 = \frac{8!}{0! \times 8!} \times \left(\frac{1}{2}\right)^8 \\ &= \frac{1}{2^8} = \frac{1}{256} = 0.0039 \end{aligned}$$

となります。この確率は、全部で 8 個中 + が 8 個、偶然生じる確率を求めています。これは、コインを 8 回投げて偶然 8 回すべて表が(もしくはすべて裏が)出る確率に等しい状況に対応します。このままでは、片側検定ですので、この数値を 2 倍して両側検定を行ないます。2 倍しても、 $2p_0 = 2 \times 0.0039 = 0.0078$ と極端に小さな値です。したがって、PTMC 前後の僧帽弁口面積の差は有意であり、PTMC は統計学的に有効なものと考えてよいこととなります。

また、マクネマー検定では、

$$\chi_0^2 = \frac{(8-0-1)^2}{8+0} = \frac{49}{8} = 6.125$$

となります。この χ_0^2 の値は、自由度 1 のカイ 2 乗分布に従いますので、有意確率 $p = 0.013$ と求めることができます。検定の結論は符号検定と変わりませんが、p 値はかなり異なっていることがわかります。ただし、実際上の問題として、帰無仮説が棄却できないというわけではありません。この

例では、特に極めて小さな p 値であり、検定を行なう場合に両者の相違は実質的にはたいした問題にはならないといえるでしょう。

ウィルコクソンの符号付順位和検定

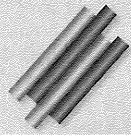


表2のデータについて、単に手術前後の僧帽弁口面積の大小をもとに符号検定を行なうのでは、すべての有用な情報を使用しているとはいえません。処理の前後の差の数値自体は用いないけれども、その符号と数値の大きさによる順位の情報を用いて、より効率よく検定する方法があります。

いま、 N 組のペアのデータがあるものとします。まず、各ペアでの測定値の差を求めます。次に、これらの各ペアの差の値について、符号を無視して、小さい順に順位を付けます(大きい順に番号を付けてもよい)。最後に、差のデータの符号が負のペアには、その順位にも負の符号を付けることにします。なお、ここでは差に0のペアはなく、また差の値(絶対値をとった場合にも)が等しいもの(タイ)はないものとします。差が0やタイのある場合については後述します。

上記の手順により、正の符号をもつ順位が、 $s_1 < s_2 < \dots < s_n$ と n 個あり、負の符号をもつ順位は、絶対値をとったものが、 $r_1 < r_2 < \dots < r_m$ と m 個あるものとします。

正の符号の順位和を S 、負の符号の順位和を R とします。ここで、 S と R の和は1から N までの総和に等しいので、

$$S + R = \frac{N(N+1)}{2}$$

の関係が成り立ちます。なお、上記の説明でわかるように、 S と R を入れ換えることが可能なので、実際の計算では差をとる方向は任意でよく、分析においては考えやすい方向に決めればよいことになります。

正の符号の順位和が負の符号の順位和よりも大きい、すなわち、 $S > R$ とすると、検定の方法は、

$$S \geq c_1 \quad \text{または} \quad R \leq c_2$$

のとき、有意差ありとすればよいことになります。ここで、 c_1 と c_2 は定数であり、通常は、この検定のために用意された数値表を用いる必要があります。統計数値表やノンパラメトリック検定の記述のある統計学の書籍にはたいてい掲載されています。しかし、標本数が大きい場合や、差にタイや0のある場合は数値表には掲載されていないので、自分で正確な有意確率を求めるか、後述するように正規近似を用いて検定をする必要があります。

差に0のある場合、最も簡単な処理の方法は、符号検定と同様にそれを分析から除外してしまうことです。その後の手順は、これまでの説明の通りです。ただし、全標本数 N は正の符号の順位の数 n 、負の数 m とした場合、新たに、 $N = n + m$ として検定を行ないます。

差にタイのある場合の扱いは、タイとなっているものに、以下に説明するようにその中間順位を与えればよいです。ただし、ここでは差の絶対値について、タイの有無を調べる点に注意が必要であり、符号を付けるのは、各ペアの差に中間順位を与えた後に行ないます。

全体のデータのうち、差の絶対値が異なる値は k 個あるとします。そのうち最小値は t_1 個、次に大きい値は t_2 個、以下順に最大値は t_k 個あるものとします。i 番目に大きな値をとるデータには、その順位として s_i を与えるものとします。すなわち、

$$s_i = \sum_{j=1}^{i-1} t_j + \frac{1+t_i}{2}$$

のように順位をつけます ($i=1, 2, \dots, k$)。この中間順位に符号を付け、正の順位和 S^* と負の順位和 R^* を求めます。

検定の考え方は、タイのない場合と同様です。しかし、タイのある場合について、一般に使用できるような検定用の数値表は用意されていませ

ん。これは、タイのある場合について、一般化が不可能なためです。このため、正確な有意確率を求めるためには、どちらかの順位和について、それ以上またはそれ以下の場合の数をすべて数え上げる必要があります。

ここで、可能な順位和の全体的場合の数は、各順位について符号の付け方が正負の2通りあるので、 2^N 個あることとなります。標本数が小さい場合、場合の数を求めるのはそれほど困難ではありませんが、標本数が大きくなるにつれ、数え上げは極めて難しくなるでしょう。そのような場合、正規分布による近似計算を用いるほうが実際的です。すなわち、

$$z_0 = \frac{|S^* - N(N+1)/4|}{D}$$

ここで、

$$D = \sqrt{N(N+1)(2N+1)/24 - \sum_{i=1}^k (t_i^3 - t_i)/48}$$

です。この z_0 が標準正規分布に従うことを利用して検定を行ないます。すなわち、両側5%点1.96、もしくは1%点2.58よりも z_0 が大きければ有意差ありとするか、正規近似により有意確率を求めればよいでしょう。

タイのない場合には、 $t_1=t_2=\dots=t_k=1$ として計算すればよく、上の D の計算式の2番目の項は0になります。なお、タイのない場合に正規近似を用いるには、 $S > R$ の場合、連続性の補正のために、 S から0.5を引いて、計算するほうがよいでしょう(逆に、 $S < R$ ならば0.5を加えます)。

【例題3】

表2のデータをもとに、符号付順位和検定を行なってみましょう。

表2のデータから、計算に必要な各データの符号付順位を求めて、表3に示しました。ここでは、すべてのペアが正の順位となっているので、正の

表3 符号付順位和検定のための表

前後の差	差の絶対値の順位	符号付順位
0.5	6	+6
0.9	8	+8
0.5	6	+6
0.5	6	+6
0.4	3.5	+3.5
0.4	3.5	+3.5
0.1	1.5	+1.5
0.1	1.5	+1.5

順位和 S^* 、負の順位和 R^* は、それぞれ、

$$S^* = 8 \times (8+1)/2 = 36$$

および、 $R^* = 0$ となります。

標本数8で、正の順位和が36となる確率を求めればよいのですが、これは負の順位和が0となる確率と等しくなります。それぞれのペアが+か-の2通りですので、全体では場合の数は、 $2^8 = 256$ 通りあります。負の順位和が0になるのは1通りしかないなので、正確な確率は、 $p_0 = 1/2^8 = 0.0039$ となりますが、これは片側確率なので2倍にして、 $p = 2 \times p_0 = 2 \times 0.0039 = 0.0078$ となります。この有意確率は、符合検定の場合と同じ結果になっています。

正規分布近似を用いて有意確率 p を求めてみましょう。まず、

$$z_0 \text{ の分子} = |36 - 8 \times (8+1)/4| = 18$$

および、

$$\begin{aligned} z_0 \text{ の分母} &= \frac{\sqrt{8 \times (8+1) \times (2 \times 8+1)/24 - [(2^3-2) + (2^3-2) + (3^3-3)]/48}}{\sqrt{50.25}} \\ &= \sqrt{50.25} \end{aligned}$$

となり、

$$z_0 = \frac{18}{\sqrt{50.25}} = 2.539$$

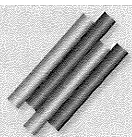
と求められます。 z_0 の値は2.539ですから、標準正規分布の両側5%点1.96よりも大きく、有意差

があることとなります。なお、 z_0 の値から正規分布近似で p 値を求めると0.011となり、やはり有意であることがわかります。

ここまで、同一の例題について異なる3種類の検定法を用いてみましたが、 t 検定による方法で有意確率が小さくなっていました。ここでは、符号検定とウィルコクソンの符号付順位和検定は、同一の検定結果を示しました。これは、8個のデータの前後の差がすべて正になっていたという特別なケースだったことによります。一般には、符号検定はデータの順序関係の情報を用いないため、他の2つの検定に比べ効率が悪いことが知られています。データの数値自体に意味がある場合には、変数の母集団での分布として正規分布でなくとも、多峰性分布でなければ中心極限定理から統計量の分布に正規性を仮定できるので、 t 検定を用いるのが最良の選択になります。

数値に絶対的な意味がなく、分布にどのような分布も仮定したくなければ、符号付順位和検定が適していることとなります。符号検定は、数値もその順位も利用できない場合に有効な手法といえます。なお、マクネマー検定は近似的な方法であり、符号検定は2項分布による確率計算を、より正確に行なっていることに注意が必要です。

対応のあるデータからなる2標本の比較



対応のあるデータを用いることで、繰り返し測定の間に行なわれた介入や特定の処理の効果を検討することができますが、その時間経過中に何か他の要因の影響により、あたかもその介入の影響のような結果が得られたのではないかという疑義が生じる可能性は否定できません。

この問題に対処するためには、介入を行なうケースのみではなく、何も行なわない、もしくは標準的な処理を行なうようなグループを設定し、比較のためのコントロールとすればよいでしょう。

表4 2グループの前後の尿酸値(mg/dL)

A薬群		標準薬群	
前	後	前	後
11.0	6.5	13.5	7.5
10.2	8.0	10.7	6.6
9.8	7.5	11.2	9.3
10.9	8.1	13.1	6.8
9.0	7.2	10.4	5.9
9.7	8.9	9.6	6.8

以下のような例題を考えてみましょう。

【例題4】

高尿酸血症患者に対して、無作為に6例にA薬を、6例に標準薬をそれぞれ投与したとします。投与前と投与1週間後の尿酸値を調べたところ、表4のようなデータが得られたとします。このデータから、A薬は標準薬に比べて効果があるといえるでしょうか。

検定の方法は、以下のようにすれば簡単に行なうことができます。

2グループの各標本についてのデータは、投与前後の対応のあるものとなっています。そこで各薬剤の効果は、投与前後の測定値の差で示せるものと考えられます。したがって、まず表5のように測定値の差を各標本について求めます。なお、表5では投与前の値から投与後の値を引いたものを示しています。

表5のような投与前後の差のデータに変えることにより、すでに連載3回目(本誌47巻3号)で取り上げた「2つの母平均値の差の検定」の問題として扱えるようになります。実際に計算してみましょう。

まず、A薬群の平均値 \bar{x}_A と不偏分散 s_A^2 を求めます。表5に示した前後の差の総和と2乗和から、

$$\bar{x}_A = 14.4/6 = 2.4$$

および、

表5 2グループの尿酸値の前後の差

A薬群	標準薬群
4.5	6.0
2.2	4.1
2.3	1.9
2.8	6.3
1.8	4.5
0.8	2.8

$$s_A^2 = 7.54/5 = 1.508$$

のように求められます。

同様に、標準薬群の平均値 \bar{x}_C と不偏分散 s_C^2 を求めます。計算すると、

$$\bar{x}_C = 25.6/6 = 4.267$$

および、

$$s_C^2 = 14.973/5 = 2.995$$

と求められます。

すでに、説明したように「2つの母平均値の差の検定」は2群の母分散が等しいか否かにより用いる方法が異なります。まず「等分散の検定」(F検定)を行いません。この検定は、2つの不偏分散の比を計算するだけです。すなわち、

$$F_0 = \frac{s_C^2}{s_A^2} = \frac{2.995}{1.508} = 1.986$$

を求めます。この F_0 は自由度(5,5)のF分布に従いますので、その上側確率を求めると、 $p_0 = 0.235$ となります。両側確率にすると5%を大きく上回っていますので、等分散を仮定することができるものと考えてよいでしょう。したがって、ここでは等分散の場合の2つの母平均値の検定を行いません。

母分散が等しい場合の帰無仮説「2グループの母平均値 μ_A , μ_C は等しい; $\mu_A = \mu_C$ 」の検定統計量は、

$$t_0 = \frac{|\bar{x}_A - \bar{x}_C|}{\sqrt{\frac{(6-1) \times s_A^2 + (6-1) \times s_C^2}{6+6-2} \times \left(\frac{1}{6} + \frac{1}{6}\right)}} = \frac{|2.4 - 4.267|}{\sqrt{\frac{5 \times 1.508 + 5 \times 2.995}{10} \times \frac{12}{36}}} = 2.155$$

と求められます。

この統計量 t_0 は、自由度10のt分布に従いますので、その有意確率は $p = 0.057$ と求められます。したがって、5%水準で有意差は認められませんので、A薬は標準薬に比べて尿酸値がより減少させる効果があるとはいえないことになります。

表4に示したような対応のあるデータが2組ある場合には、表5のような差のデータの形式にすることで、通常の独立な2標本の母平均値の差を検定する問題として扱えるようになります。これは、2つの母平均 μ_1 と μ_2 について帰無仮説「 $\mu_1 = \mu_2$ 」を検定する問題と、処理の効果についての差の平均値 δ_1 と δ_2 に関する帰無仮説「 $\delta_1 = \delta_2$ 」を検定する問題が、形式的に同一であることによります。

対応のあるデータでの標本数の設定



ここでは研究計画の段階で、対応のあるデータを用いる場合に必要とされる標本数の求め方を説明します。ここでは、①1グループでの介入や処理の前後比較を行なう場合と、②比較のためのコントロールを置く場合の2標本での比較を行なう場合の2つの場合について、以下に説明したいと思います。ただし、共通する考え方について、まず説明します。

対応のある場合でも、考え方としては1ケースが介入や処理の前後に2グループあるものと考えられます。そして2グループの母平均値の差の検定が有意となるように、標本数を決めるというように考えます。標本数の決定のためには、以下の情報が必要となります。

- (1) 検定のための第1種の誤り 2α (有意水準)
- (2) 検定のための第2種の誤り β , または検出力 $1 - \beta$
- (3) 2グループの母平均値の推定値, または母平均値の差の推定値 Δ
- (4) 2グループの母分散の推定値 σ^2 , または差の母分散の推定値 σ^2

通常は、 2α は0.05(5%)とし、 β は 2α の4倍とします。したがって、有意水準が5%ならば、検出力は80%に設定するのが普通です。あと必要な情報は、母平均値の差の推定値 Δ とその母分散の推定値 σ^2 ということになります。対応のある場合のデータでは、処理の前後の測定値の差を計算するために、一般にデータそのものの母分散よりも、差の母分散の推定値 σ^2 は小さくなります。このため小さな標本数でも、効率のよい検定が行えることとなります。

2つの変量の比、 Δ/σ は標準化効果量と呼ばれることがあります。この値がわからないと、適切な標本数を求めることはできません。

さて、標準正規分布の上側 α 点の値を z_α とします。2グループの全体の標本数を N とし、2グループのそれぞれの標本数を $n_1 = pN$, $n_2 = qN$, $p+q=1$, とします。このとき、全体の標本数は、

$$N = (1/p + 1/q) \times (z_\alpha + z_\beta)^2 \times \frac{\sigma^2}{\Delta^2} + \frac{z_\alpha^2}{4}$$

によって近似的に求められます。

有意水準を両側5%、検出力を80%とすると、標準正規分布の上側2.5%点は $z_{0.025} = 1.96$ 、また上側20%点は $z_{0.2} = 0.842$ となります。1グループで対応のある場合には、2グループの標本数 n を等しいものとし、

$$n = 15.7 \times \frac{\sigma^2}{\Delta^2} + 0.5$$

と簡単に、標本数を求めることができます。

【例題5】

表2のPTMC前後の僧房弁口面積のデータで、検定結果は極めて高度に有意でしたが、実際にはどのくらいの標本数があれば5%水準で有意な結果が得られたのか、検討してみたいと思います。

まず、差の平均値は $\bar{d} = 0.424$ 、差の母分散の推定値である不偏分散は $s^2 = 0.065$ でしたので、有意水準5%、検出力80%として、これらの数値を標本数の計算式に代入してみます。

$$n = 15.7 \times \frac{0.065}{0.424^2} + 0.5 = 6.18$$

となりますので、7例あれば十分だったことがわかります。

介入や処理の前後で測定を繰り返す必要がありますが、このように対応のある場合には、差の母分散が小さいものと推定されるため、必要とされる標本数も小さくなるのが一般的です。

【例題6】

表5の投与前後での2グループでの尿酸値の差のデータで、実際にはどのくらいの標本数があれば5%水準で有意な結果が得られたのか、上記と同様に検討してみたいと思います。

計算に使用する計算式は同じになります。まず、2グループの平均値の差は、

$$\bar{d} = 4.267 - 2.4 = 1.867$$

となり、母分散の推定値は、

$$s^2 = \frac{5 \times 1.508 + 5 \times 2.995}{10} = 2.2515$$

と計算されます。

有意水準5%、検出力80%として、これらの数値を標本数の計算式に代入してみます。

$$n = 15.7 \times \frac{2.2515}{1.867^2} + 0.5 = 10.64$$

となります。研究中の変動を考慮して検出力を80%にすると、次に同様の研究を行なう場合には、各グループ11例ずつの研究計画にする必要があります。

おわりに



本連載「統計学のキー・ポイント—『検定』に焦点を当てて」も、今回で最終回となりました。統計学で最も必要であるのに、最も誤解しやすい検定を理解しやすく平易に解説し、必ず最後には研究計画で必要とされる標本数、調査対象者数はどのくらい必要なかを推定するための方法を説明するという方針で連載しました。実際には平易な解説は難しかったので、理解しにくい箇所がかなり残ってしまったのではないかと反省しています。

ただ、この連載を機会に、研究に必要な標本数をあらかじめ研究計画時に求めておかなければならない、という理解が広まればよいと思っています。これからも、何かまた別の話題で誌面でお会

いできればと思います。

●文献

- 高木廣文(1986). 臨床研究のための統計学—IV. 特殊な形式のデータの検定. 臨床泌尿器, 40(10), 797-801.
- 高木廣文(1999). 対応のあるデータの比較. 超音波検査技術, 24(6), 424-429.
- 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.138-140.
- 高木廣文(2009). ナースのための統計学, 第2版. 医学書院. pp.88-90.
- 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.
- 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2回 検定と標本数の関係. 看護研究, 47(2), 150-154.
- 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3回 平均値の差についての検定. 看護研究, 47(3), 264-269.
- 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4回 割合についての検定. 看護研究, 47(5), 474-481.
- Hulley, S.B., Cummings, S.R., Browner, W.S., Grady, D.G., & Newman, T.B.(2007). *Designing Clinical Research*(3rd ed.). Lippincott Williams & Wilkins. / 木原雅子, 木原正博訳(2009). 医学的研究のデザイン, 第3版. メディカル・サイエンス・インターナショナル, pp.67-97.

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

トラブルに巻き込まれないための医事法の知識

著 福永篤志
法律監修 稲葉一人

すべての医療人に向けた、医療紛争に巻き込まれないために知っておくべき法律知識の解説書。臨床医の目線で日常診療上注意すべき法律50項目を選び、具体的な判例を交え、1項目につき3ページ程度で分かりやすく噛み砕いて解説。「medicina」「脳神経外科」誌の好評連載コラム「医事法の扉」の単行本化。

●B6 頁344 2014年
定価: 本体2,200円+税
[ISBN978-4-260-02011-4]