

統計講座 (第6回)

対応のあるデータの比較

高木 廣文

はじめに

高血圧症の患者に減塩させるための指導を行ったとしよう。そのような場合、減塩指導が本当に効果があったかを、どのように調べればよいだろうか。単純に考えれば、指導を受けた患者について、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、実施した指導が有効であったと考えてよいことになるだろう。同様に、ダイエットの効果を知りたければ、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だろう。また、学生に講義を行い、その講義が有効であったかは、講義内容に関連するような知識量の増加がひとつの目安になるだろう。そのような場合には、同一試験を講義の前後に行うことによって、得点の増加を調べることで、講義の評価が行えるだろう。

このように、ある対象に対して特定の介入を行った場合、その効果は適当な項目を繰り返し測定し、そのデータの差を比較することにより、実施した介入の評価を行うことができる。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ており、通常は「対応のあるデータ」と呼ばれることが多い。

対応のあるデータとしては、上記のような摂取量や体重のような量的な場合や、ある臨床症状の有無や軽減などのように質的な場合もある。これらの対応のあるデータを扱う方法は、図1に示したように、主に3つの方法がよく使われている。

それらについて、以下に順に説明することにしよう。

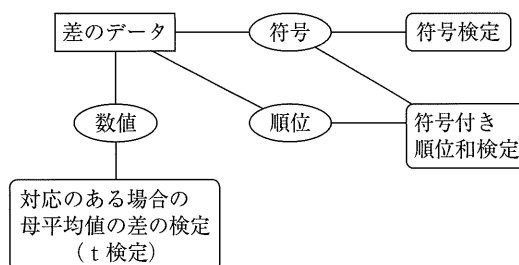


図1 対応のあるデータの検定方法

対応のある量的データの場合

表1は僧帽弁狭窄症患者20名について、PTMC（経皮経管的僧帽弁拡張術）を行う前後の僧帽弁口面積を心エコーにより求めたデータである（模擬データ）。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だろうか。

個人について同一の調査や測定を繰り返して行う場合、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」と呼ばれる方法である。

以下に簡単に、理論的な解説をしよう。

いま、ケース数を N とし、指導などの介入（処理）の前後での変数の測定値を、それぞれ (x_1, y_1) , (x_2, y_2) , ..., (x_N, y_N) としよう。ここで、処理や介入としては「ある薬を服用させる」、「手術をする」、「食事指導を行う」などがあるだろう。これらの効果は、その介入の前後での測定値の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る。

介入前後の測定値の差を

表1 指導の前後での僧帽弁口面積と差

No	前	後	差
1	1	1.5	0.5
2	1.1	2	0.9
3	0.6	1.1	0.5
4	0.5	1	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
9	1.5	2.6	1.1
10	0.6	0.9	0.3
11	1.4	1.5	0.1
12	0.4	0.9	0.5
13	0.6	1	0.4
14	1	1.6	0.6
15	0.6	1	0.4
16	0.5	1	0.5
17	1.1	1.9	0.8
18	1.4	2.4	1
19	1.8	1.9	0.1
20	0.5	1.2	0.7

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

としたとき、差の平均値 d_M 、および差の不偏分散 s^2 をまず求める。

検定の帰無仮説は、「母集団での差の母平均値は0」である。検定のための統計量として、

$$t_0 = \frac{|d_M| \sqrt{N}}{s}$$

を求める。上記の統計量 t_0 は、自由度 $N-1$ の t 分布に従うことで仮説の検定が行える。

すなわち、仮説の検定を行うには、 t 分布表より両側 α 点（5%か1%）の値 $t_{\alpha/2}$ を求め、計算して求めた t_0 と比較し、

$$\begin{aligned} t_0 \geq t_{\alpha/2} &\rightarrow \text{水準}\alpha\text{で有意差あり} \\ t_0 < t_{\alpha/2} &\rightarrow \text{水準}\alpha\text{で有意差なし} \end{aligned}$$

のように検定を行えばよい。

上記の手順で検定を行い、有意差が認められた場合、行った処理に効果があったものと考えてよいことになる。

なお、この検定では、分析を行う変数の分布が、治療や指導などの介入により、値の大きい方向（もしくは小さい方向）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。したがって、処理の前後の測定値間には相関が認められるはずである。もしも測定値間に相関が認められない場合、ここで紹介した検定方法は適切なものではないことになる。

実際に計算する場合には、表1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る。次に、差の平均 d_M と差の不偏分散 s^2 を求める。これらの値は、

$$d_M = (0.5 + 0.9 + \dots + 0.7) / 20 = 0.495$$

$$\begin{aligned} s^2 &= [(0.5 - 0.495)^2 + \dots + (0.7 - 0.495)^2] / 19 \\ &= 0.08579 \end{aligned}$$

となる。したがって、検定のための統計量は、

$$t_0 = \frac{0.495 \times \sqrt{20}}{\sqrt{0.08579}} = 7.559$$

と t_0 の値が求められる。この t_0 は t 分布に従い、その自由度は、 $20 - 1 = 19$ となる。

統計数値表から自由度19の t 分布の両側5%点と1%点を求めると、それぞれ2.093, 2.861である。 t_0 の値はこれらより大きいので、1%水準で有意差が認められるものと考えてよい（ p 値を計算すると、0.000001未満と極めて小さい）。したがって、PTMCの僧帽弁口面積増加については有効なものと考えてよいことになるだろう。

ちなみに、カテーテル手術前後の僧帽弁口面積の単相関係数を求めると、0.824となり、かなり高い相関を示している。したがって、全体的に数値が一定方向に変化したものと考えてよいことになり、本方法による検定は妥当なものと考えてよいだろう。

なお、ここで紹介した方法は、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ったが、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（もしくは、曝露-対照研究）などでも有効な方法である。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健常者などの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比較することも可能である。

■ 符号検定による比較

ここでは、表1のデータの差の値をそのまま使用するのではなく、ノンパラメトリックに検定する方法を説明しよう。

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法を用いる利点は、母集団で正規分布を仮定することに無理がある場合や、データの数値にそれほど信頼がおけないが、大小関係程度の情報は利用できる場合などでも適用が可能な点にある。

いま、ペアを組んでいる一方のグループをAとし、他方をBとしよう。その処理の効果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の方が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よい。

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「-」としよう。また、どちらも同一の結果（0の「タイのある場合」と呼ぶ）を示していれば、それは検定のための情報がないものとして以後の分析から除外する。

正（+）を示した対の数を N_+ とし、負（-）を示した対の数を N_- としよう。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、 $N = N_+ + N_-$ となる。

特定の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 $1/2$ となるはずである。

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、つぎのような2項分布により求められる。すなわち、

$$\begin{aligned} p_0 &= \sum_{k=N_+}^N \frac{N!}{(N-k)! \times k!} \times (1/2)^N \\ &= \sum_{k=0}^{N_-} \frac{N!}{(N-k)! \times k!} \times (1/2)^N \end{aligned}$$

により求められる。上記の関係から、測定値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だろう。

ここで求められた確率は、一定方向への偏りについて計算されている。したがって、事前に何らかの情報により、一方方向にのみ結果が偏ることが明らかな場合には、 p_0 の値が0.05以下（5%）もしくは0.01以下（1%）の場合に有意差ありとすればよい（片側検定）。しかし、通常は確実な事前情報がない場合が大部分であり、また差の有無のみを問題とし、かつ検定結果に安全を見込んで、 p_0 を2倍した値を、検定の有意確率として用いる（両側検定）ことが多い。

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っている。そのため、数値自体を用いる t 検定に比べ、一般に検定の効率はやくない。

もとのデータが本質的に質的なものでない場合、特別な理由がない限り、後述のウィルコクソン（Wilcoxon）の符号付順位和検定か、前述の t 検定を用いる方がよいだろう。

ところで、標本数 N が大きくなると、上記の確率計算はやや煩雑になる。そのような場合、標準正規分布による近似により、検定を行うと便利である。いま、 $N_+ > N_-$ としよう。検定のためには、

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

を求めればよい。

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る。なお、上式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」を行っている。

検定の方法は、この z_0 の値が両側5%点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。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、それに基づいて行う検定は「McNemar検定（マクネマー検定）」と呼ばれている。

表1では、20対のデータの差はすべて+なので、計算は20個中20個が正になる確率を求めればよい。

$$p_0 = \sum_{k=20}^{20} \frac{20!}{(20-k)! \times k!} \times (1/2)^{20}$$

$$= 1/2^{20} = 0.000000953$$

となる。

上記の確率は、全部で20個中+が20個、偶然生じる確率を求めている。これは、コインを20回投げて偶然20回すべて表が（もしくはすべて裏が）出る確率に等しい。片側検定でなく、両側検定を行うにはこの数値を2倍して用いればよいのだが、どちらにしても極端に小さな値である。したがって、PTMCは僧帽弁口面積の増加に統計学的に有効なものと考えてよいことになる。

正規分布を用いた検定では、

$$z_0 = \frac{2 \times 20 - 20 - 1}{\sqrt{20}} = 4.2485$$

となる。この値は、標準正規分布の両側1%点の値2.58よりも大きいので、1%水準で有意差ありという結論になる（p値は約0.00002程度）。検定の結論は変わらないが、実際のp値はかなり異なっている。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訳ではない。この例では、特に極めて小さなp値であり、両者の相違は実質的には何の問題にもならない。

■ ウィルコクソンの符号付順位和検定

表1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うのは、すべての情報を使用しているとはいえない。そこで、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の順位をデータとして、より効率よく検定する方法がある。

いま、N組のペアのデータがあるものとしよう。まず、各ペアでの測定値の差を求める。次に、これらの差の値について、符号を無視し

て、小さい順に順位を付ける（大きい順に番号を付けても同じことである）。最後に、差の値の符号が負のものには、その順位にも負の符号を付けることにする。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）はないものとする。そのような、0やタイのある場合の説明については後述する。

上記のような手順により、正の符号を持つ順位が、 $s_1 < s_2 < \dots < s_n$ 、とn個あり、負の符号を持つ順位は、絶対値をとったものが、 $r_1 < r_2 < \dots < r_m$ 、とm個あるものとしよう。

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

$$S + R = N(N+1)/2$$

の関係が成り立つ。ここまでの説明で、SとRを入れ換えることが可能なことから、差をとる方向は任意でよく、分析において考えやすい方向に決めればよいことがわかるだろう。

検定の方法として、正の符号の順位和の方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、

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

のとき、有意差ありとすればよい。なお、 c_1 と c_2 は定数であり、検定のためには、符号付順位和検定のための数値表を用いればよい。通常の統計数値表や、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大抵掲載されているはずである。ただし、標本数が大きい場合や、差にタイや0のある場合は、観測された順位和と同一になる場合の数を自分でデータの組み合わせから数え上げ、有意確率を求めるか、後述する正規近似による検定を行わねばならないので、やや煩雑になる。

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ある。その後の手順は、これまでの説明のとおりである。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 $N = n + m$ として検定する。

差にタイのある場合の扱いは、すでに説明したウィルコクソンの順位和検定と同様に、タイとなっているものにその中間順位を与えればよい。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して欲しい。符号を付けるのは、中間順位を与えた後である。

全体のデータのうち、差の絶対値が異なる値は k 個あるものとしよう。そのうち最も小さい値は t_1 個、次に大きい値は t_2 個、以下順に最大のものが t_k 個あるものとする。 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 を与えるものとしよう。すなわち、

$$s_i = (1 + t_i) / 2 + \sum_{j < i} t_j,$$

のように順位を付ければよい ($i=1, 2, \dots, k$)。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^* を求める。

検定の考え方は、タイのない場合と同様であるが、タイのある場合について、一般に用いることが可能な検定用の数値表は用意されていない。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ある。そのため、有意確率を求めるには、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る。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 個あることになる。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ないが、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。そのような場合、正規分布による近似計算を用いる方が实际的である。

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

$$D = \sqrt{[N(N+1)(2N+1)/24 - \sum_i (t_i^3 - t_i)/48]}$$

を求め、 z_0 が標準正規分布に従うことを利用すればよい。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

すればよい。

タイのない場合には、 $t_1 = t_2 = \dots = t_N = 1$ として、計算すればよく、上式の右辺の2番目の項は0になる。ただし、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、計算する方がよいだろう (逆に、 $S < R$ ならば0.5を加える)。

表2 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	13	+13
0.9	17	+17
0.5	13	+13
0.5	13	+13
0.4	8.5	+8.5
0.4	8.5	+8.5
0.1	2.5	+2.5
0.1	2.5	+2.5
1.1	20	+20
0.3	6	+6
0.1	2.5	+2.5
0.5	13	+13
0.4	8.5	+8.5
0.6	16	+16
0.4	8.5	+8.5
0.5	13	+13
0.8	18	+18
1	19	+19
0.1	2.5	+2.5
0.7	17	+17

(この例では、すべての差のデータが正なので、絶対値と符号付順位は等しい)

実際の計算を行ってみよう。表2に各データの符号付順位を示しておいた。ここでは、すべてのペアが正の順位となっているので、正の順位和を S^* 、負の順位和 R^* は、それぞれ、

$$S^* = 20 \times (1 + 20) / 2 = 210, \quad R^* = 0$$

となる。

標本数20で、正の順位和が210となる確率を求めればよいが、これは負の順位和が0となる確率と等しい。この例は、データに「タイのある場合」なので、正規分布を用いて有意確率 p を求めよう。

$$\text{分子} = |210 - 20 \times (20 + 1) / 4| = 105$$

$$\begin{aligned} \text{分母} &= \sqrt{[20 \times (20 + 1) \times (2 \times 20 + 1) / 24 - \{(4^3 - 4) \\ &\quad + (4^3 - 4) + (5^3 - 5)\} / 48]} \\ &= \sqrt{719.125} \end{aligned}$$

となる。したがって、

$$z_0 = 105 / \sqrt{719.125} = 3.915$$

と求められる。

z_0 の値、3.915は標準正規分布の両側5%点1.96よりも大きく、有意差があることになる（両側確率は、約0.009%に相当する）。

■ おわりに

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た。結果をみると、t検定によるものとウィルコクソンの符号付順位和検定の結果は、t検定がやや有意確率が小さくなっている。ここでは、符号検定もかなり効率のよい検定結果を示している。これは、データが極めて特別なせいである。すなわち、20個のデータの前後の差がすべて正になっていたことに

よる。

本例ではそうではなかったが、一般的にいうと、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ようである。

データの数値自体に意味があり、変数の母集団での分布として正規分布を仮定することに無理がなければ、t検定を用いるべきである。しかし、数値に絶対的な意味がなく、分布に正規性を仮定するのも無理があれば、符号付順位和検定が適しているだろう。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る。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しよう。

今回の方法も、私のホームページで簡単に計算できるので、試してみるとよいだろう。

参考文献

- 1) 統計数値表編集委員会編：簡約統計数値表，日本規格協会，1977.
- 2) 高木廣文，三宅由子：よくわかる医療・看護のための統計学入門—パソコンによるデータ解析，メディカ出版（大阪），1991.