

●講座/臨床研究のための統計学

Ⅱ. 対応のある場合の2標本の比較*

高 木 廣 文**

対応のある場合の2標本の比較

1. 対応のある2標本の母平均値の差の検定

【例題 2・1】

頻尿患者10例について、A薬の投与前と1週間後の1日あたりの排尿回数を調べたところ、第1表のようになった。このデータから、A薬の治療効果はあるといえるか。

第1表 薬物投与前と1週間後の排尿回数

患 者	前	後	前後の差
1	20	12	8
2	13	8	5
3	9	7	2
4	12	14	-2
5	15	13	2
6	9	10	-1
7	18	10	8
8	14	9	5
9	25	17	8
10	16	11	5

【検定法 2・1】

ここでは、排尿回数の母集団での分布が正規分布で近似できるものとして、t検定を用いることにする。

第1表の最右欄に示したように、まず、投与前後での排尿回数の差を求める。差の平均 $\bar{d}$ と差の不偏分散 s^2 は(2.1)式より、

$$\bar{d} = (8+5+\cdots+5)/10=4$$

$$s^2 = \{(8-4)^2 + \cdots + (5-4)^2\}/9 = 13.333$$

となる。これを、(2.2)式に代入すると、

$$t_0 = 4/\sqrt{(13.333/10)} = 3.464$$

と求められる。 t_0 は自由度 $(10-1)=9$ のt分布に従う。統計数値表から自由度9のt分布の両側5%点と1%点を求めると、それぞれ2.262, 3.250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。

上記の結果から、A薬による治療効果は有効なものと考えてよいことになるだろう。

【解説 2・1】

標本数を n とし、ある処理の前後での変数の測定値をそれぞれ $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$ とする。なお、これらの変数は母集団では、独立に正規分布に従っているものと仮定する。また、ある処理とは「特定の薬物を服用させる」、「手術をする」、「食事指導を行う」などの場合が考えられる。これらの処理の効果は、処理の前後での測定値の差があるか否かにより、その有無を判断できるだろう。

処理の前後の測定値の差を $d_i = x_i - y_i$ ($i=1, 2, \dots, n$)としたとき、差の平均値 $\bar{d}$ 、および差の不偏分散 s^2 は、

$$\bar{d} = \frac{1}{n} \sum_{i=1}^n d_i, \quad s^2 = \frac{1}{n-1} \sum_{i=1}^n (d_i - \bar{d})^2 \quad (2.1)$$

により求められる。

仮説として、「母集団での差がない」とする。この仮説のもとで、

$$t_0 = \frac{|\bar{d}|}{\sqrt{(s^2/n)}} \quad (2.2)$$

は、t分布に従うことで仮説の検定が行える。ただし、 t_0 の自由度 ν は、

* Statistics in Clinical Studies: Ⅱ. Methods of the Comparison for the Matched Pair Samples

** 聖路加看護大学 Hirofumi Takagi: St. Luke's College of Nursing

$$\nu = n - 1 \quad (2.3)$$

である。仮説の検定には、 t 分布表より両側 α 点 (5% か 1%) の値 $t_{\alpha/2}(\nu)$ を求め、 t_0 と比較する。

$$t_0 \geq t_{\alpha/2}(\nu)$$

ならば、水準 α で有意差があるものとし、そうでなければ、差があるとはいえない。

上記の検定により有意差が認められる場合、行つた処理に特定の効果があつたものと考えてよいことになるだろう。

なお、この検定では、特定の変数の分布がある処理により、値の大きい方向 (もしくは小さい方向) に全体として「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。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くなる。ちなみに、第1表のA薬投与前後の排尿回数の単相関係数を求めると、0.680となり、かなり高い相関を示している (5%水準で有意)。

また、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つたが、患者一対照研究 (もしくは、曝露一対照研究) などでも有効な方法である。すなわち、性・年齢・職業などの条件をマッチさせた者同士をペアとして各標本を設定し、マッチさせなかつた各種の特性について比較することも可能である。

2. 符号検定による比較

【例題 2・2】

例題 2・1 の頻尿患者のデータについて、排尿回数の差の値自体にあまり意味がなく、かつ母集団で正規分布を仮定することに無理があるものとして、A薬の効果があつたかないかを調べてみよう。

第 2 表 符号検定のための表

前後の差	8	5	2	-2	2	-1	8	5	8	5
符号	+	+	+	-	+	-	+	+	+	+

【検定法 2・2】

ここでは、A薬投与前後の排尿回数の増減のみ

に着目するとしよう。投与前の排尿回数が投与後より大きい場合に「+」、小さい場合に「-」として第2表に示した。各符号の個数は、正が8個に負が2個となる。治療効果がないと仮定した場合、正か負かは確率 1/2 と等しいはずである。したがって、10 例中 8 例以上が正となる確率 p_0 は、後述の (2.5) 式の 2 項分布を用いて、

$$\begin{aligned} p_0 &= \sum_{k=8}^{10} \frac{10!}{(10-k)! \cdot k!} (1/2)^{10} \\ &= \sum_{k=0}^2 \frac{10!}{(10-k)! \cdot k!} (1/2)^{10} \\ &= (1 + 10 + \frac{10 \times 9}{2}) \times (1/2)^{10} \\ &\div 0.05469 \end{aligned}$$

となる。上記の確率は、正が8個以上、もしくは負が2個以下の場合についてのものである (片側検定)。治療効果がどのような傾向をもつか、事前に断定できない場合には、 p_0 を2倍した値、すなわち $2p_0$ を求めるべきである (両側検定)。その値は、約 0.109 となる。したがって、符号検定では 5%水準で有意とならず、A薬の治療効果があるとはいえない。

【解説 2・2】

いま、 n 対の標本があるものとしよう。すでに述べたように、各対は患者一対照のように疾患の有無によつてペアを作つてもよいし、一方にある治療薬を投与し、他方に偽薬を投与するような組合せでもよい。また、本例のように、同一標本に特定の処理をし、その前後での測定値をデータとしたものでもよい。

一方の群にある処理を施し、これを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つてもよいし、各対について、「Aの方がBより有効である」といつた質的なものでもよい。各対についてAがBより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) を示していれば、それは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、正と負の対の個数のみを用いるので、検定の対象となる全標本数を新たに N とす

ると、

$$N = n_+ + n_- \quad (2.4)$$

となる。処理の効果がなければ、各対で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 p_0 は、N 個中 n_- 個以下が負となる確率と一致し、単純な 2 項分布により求められる。すなわち、

$$p_0 = \sum_{k=n_+}^N \frac{N!}{(N-k)! \cdot k!} \cdot (1/2)^N$$

$$= \sum_{k=0}^{n_-} \frac{N!}{(N-k)! \cdot k!} \cdot (1/2)^N \quad (2.5)$$

により求められる。したがって、+- の符号の付け方は結果に無関係である。

上記の確率は、一定方向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、通常は確信を持つて、そのように断言できないことが多いはずである。したがって、一般には 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)。

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、検定の効率はよくない。もとのデータが質的なものでない場合、特別な理由がない限り、後述の Wilcoxon の符号付順位と検定か、前述の t 検定を用いるほうがよいだろう。

標本数 N が大きくなると、(2.5) 式による確率計算は、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。正の個数の期待値 $E(n_+)$ と分散 $V(n_+)$ は、

$$E(n_+) = N/2, \quad V(n_+) = N/4 \quad (2.6)$$

となる。これを用いることで、

$$z_0 = \frac{2n_+ - N - 1}{\sqrt{N}} \quad (2.7)$$

が求められる。 z_0 は、N 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「連続性の補正」を行っている (前回解説参照)。 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、2.58 より大きければ同様に 1% で有意差ありとすればよい。

3. Wilcoxon の符号付順位と検定

【例題 2・3】

第 1 表のデータについて、たんに A 薬投与前後の排尿回数の大小をもとに符号検定を行うのではなく、排尿回数の差の数値自体は用いないものの、その符号と大きさの順位をデータとして A 薬の効果を検定する。

第 3 表 符号付順位と検定のための表

前後の差	差の絶対値の順位	符号付順位
8	9	9
5	6	6
2	3	3
-2	3	-3
2	3	3
-1	1	-1
8	9	9
5	6	6
8	9	9
5	9	6

【検定法 2・3】

まず、排尿回数の前後の差を求めることは、これまで説明した方法と同様である。つぎに、差のデータについて、その絶対値 (符号を無視したもの) について小さい順に、順位を付ける。もしも数値が等しいものがあれば、前回説明した順位と検定と同様に、その中間順位を与えるものとする。

順位を与えた後で、各順位についてもとの差と同じ符号を付ける。これは、差が負の値の順位に「-」を付けることを意味する。このような手順により、第 3 表が作成される。

いま、正の順位和を S^* 、負の順位和を R^* とすると、

$$S^* = 9 + 6 + 3 + 3 + 9 + 6 + 9 + 6 = 51$$

$$R^* = 3 + 1 = 4$$

である。

標本数 10 で、正の順位和が 51 以上となる確率を求めればよいが、これは負の順位和が 4 以下とな

第4表 順位和が4以下となる組合せ

順位和	順位の組	場合の数
0	空集合	1
1	(1)	1
3	(3)	3
4	(1, 3)	3
合計	—	8

* 順位和が2となる場合はない。

る確率に一致する。「タイのない場合」には、通常の統計数値表に掲載されている「Wilcoxonの符号付順位和」の表にある値と、求めた順位和の大小を比べて検定すればよい。しかし、このように「タイのある場合」には、数値表を利用することはできない。ここでは、実際に負の順位和が4以下となる確率を求めることにする。

10個の標本について可能な符号の組合せの数は、

$$2^{10} = 1024$$

ある。負の順位和が4以下となる組合せの数は、第3表の「差の絶対値の順位」を用いて、符号を無視して順位の情報のみから計算すると、第4表に示したように8通りの場合がある。したがって、負の順位和が4以下となる確率 p_0 は、

$$p_0 = 8/1024 \div 0.0078125$$

である。ここでは、この値を2倍して、両側確率により検定する。 $2p_0$ の値は0.015625、すなわち約1.6%で有意差が認められることになる。この結果、A葉の治療効果があると考えてよいことになる。

【解説 2・3】

符号検定は、標本の各対についてのデータの差の符号のみを用いるものであつた。しかし、差の値についても利用可能ならば、より効率のよい検定を行うことができる。

いま、N組の標本の対があるものとしよう。まず、各対での特定の変数の測定値の差を求める。次に、これらの差について符号を無視して、小さい順に順位をつける。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の対はなく、また差の値（絶対値をとった場合にも）が等しいもの

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 とすると、

$$\left. \begin{aligned} S &= s_1 + s_2 + \dots + s_n \\ R &= r_1 + r_2 + \dots + r_m \end{aligned} \right\} \quad (2.8)$$

である。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。上記の説明で、 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、 $S > R$ とすると、

$$S \geq c_1 \text{ または } R \leq c_2 \quad (2.9)$$

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、 S が c_1 以上となる確率と R が c_2 以下の確率が等しいように定めるのが普通である。これが「Wilcoxonの符号付順位和検定」である。検定のための定数 c_1 もしくは c_2 は、符号付順位和検定のための数値表を用いればよい。前回紹介した統計数値表や、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、例題2・3で示したように、場合の数から有意確率を直接求めるか、後述する正規近似による検定を行わねばならない。

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、これまでの説明の通りである。ただし、全標本数 N は正の符号の順位の個数 n 、負の個数 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、前回説明した「Wilcoxonの順位和検定」と同様に、タイとなつてい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。

いま、タイの個数が k 個あり、それぞれのタイには $d_1, d_2, \dots, d_k$ 個のデータが含まれているものとする。ここで、

$$N = d_1 + d_2 + \dots + d_k \quad (2.10)$$

である。中間順位 t_i ($i=1, 2, \dots, k$) は、

$$t_i = (d_i + 1) / 2 + \sum_{j=1}^{i-1} d_j \quad (2.11)$$

とかける。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の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、例題2・3のように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の場合の数は、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よいだろう。

正の符号の順位和 S^* の期待値 $E(S^*)$ と、分散 $V(S^*)$ は、

$$E(S^*) = N(N+1)/4 \quad (2.12)$$

$$V(S^*) = N(N+1)(2N+1)/24 - \left\{ \sum_{i=1}^k (d_i^3 - d_i) \right\} / 48 \quad (2.13)$$

である。これを用いて、

$$z_0 = \frac{|S^* - E(S^*)|}{\sqrt{V(S^*)}} \quad (2.14)$$

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、1%点2.58よりも z_0 が大きければ、有意差ありとする。タイのない場合には、 $d_1 = d_2 = \dots = d_N = 1$ として、(2.13) 式より分散を求めればよい。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、(2.14) 式を用いた方がよいだろう（逆に、 $S < R$ ならば0.5を加える）。

例題2・3について正規近似を用いると、

$$E(R^*) = 10 \times (10+1) / 4 = 27.5$$

$$\begin{aligned} V(R^*) &= 10 \times (10+1) \times (2 \times 10+1) / 24 \\ &\quad - \{ (3^3 - 3) + (3^3 - 3) + (3^3 - 3) \} / 48 \\ &= 96.25 - 1.5 = 94.75 \end{aligned}$$

となる。したがって、

$$z_0 = |4 - 27.5| / \sqrt{94.75} \div 2.414$$

と求められる。2.414は標準正規分布の両側約1.58%に相当する。この値は、先に正確に計算した確率とほとんど一致していることがわかる。このように、標本数が10程度でも正規分布による近似は、実際の検定に十分であると考えられる。

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 Wilcoxon 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るものの、大差ないものとなつてゐる。しかし、符号検定は他の2つの検定に比べ効率が悪いようである。この理由は、符号検定が差についての符号の情報しか用いないためである。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にも無理があれば、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。最後に、符号検定の特殊なケースとして、「McNemar の検定」を説明する。

4. McNemar の検定

【例題 2・4】

第5表は、小児の膀胱尿管逆流の程度を「軽度」と「高度」に分け、44例について保存的に5年間観察した前後の結果をまとめたものである。この結果から、保存的な経過観察で逆流の程度が軽減するといえるだろうか。

第5表 膀胱尿管逆流の経過観察前後の程度

前 \ 後	後		計
	軽度	高度	
軽度	23	3	26
高度	11	7	18
計	34	10	44

第6表 McNemar の検定用の表

A \ B	—	+	計
—	a	b	a + b
+	c	d	c + d
計	a + c	b + d	n

$$n = a + b + c + d$$

【検定法 2・4】

McNemar の検定に用いるデータは、経過観察の前後で変化のみられた対についてのものだけである。軽度から高度が3例、高度から軽度が11例である。これを、(2.15)式に代入すると、

$$\chi_0^2 = \frac{(|3-11|-1)^2}{(3+11)} = 3.5$$

となる。この値と自由度1の χ^2 分布の上側5%点である3.84と比較すると、やや小さいので有意とはいえない(有意確率は約0.0614)。したがって、保存的な経過観察で膀胱尿管逆流の程度が軽度になるとは断定できないことになる。

【解説 2・4】

McNemar の検定は、患者と対照を対にしてある特性の有無の差を調べる場合や、例題のように同一標本について特定の処理の前後の状態を調べ、その処理の効果を調べるために用いられる。第6表では、標本AとBが対になつており、ある特性の有無について「+」と「-」で示している。AとBの対について、異なる組合せのもの、すなわちAが「-」でBが「+」のb対と、Aが「+」でBが「-」のc対から、

$$\chi_0^2 = \frac{(|b-c|-1)^2}{(b+c)} \quad (2.15)$$

を計算する。 χ_0^2 は自由度1の χ^2 分布に従うので、上側5%点3.84か1%点6.63と比べ、それより大ならば有意差ありとすればよい。これが、

McNemar の検定である。なお、上式は連続性の補正をしている。

ところで、この検定は符号検定と同等であることがわかる。AとBが同一のものをタイとして分析から除外し、Aが「-」でBが「+」のものをあらたに正とし、逆にAが「+」でBが「-」のものを同様に負とすると、符号検定の説明で、正の個数 n_+ はb、負の個数 n_- はc、そしてNはb+cとなる。これらの値を、正規近似の式(2.7)に代入すると、

$$z_0 = \frac{|b-c|-1}{\sqrt{b+c}} \quad (2.16)$$

となる。なお、ここでは連続性の補正を考慮して、b-cには絶対値記号を付けた。

独立に標準正規分布に従う変数の2乗和は、 χ^2 分布に従うので、上記の z_0 の2乗は自由度1の χ^2 分布に従う。これが、(2.15)式である。したがって、McNemar 検定で近似的な有意確率を求めるのではなく、標本数が小さければ、符号検定を用いて、正確な有意確率を求めることができる。

第5表の例では、14例(ペア)のうち正が3以下となる確率 P_0 は、

$$P_0 = \left(1 + 14 + \frac{14 \times 13}{2} + \frac{14 \times 13 \times 12}{3 \times 2}\right) \times (1/2)^{14} \div 0.0286865$$

である。この値を2倍すると約0.0574となる。すなわち、有意確率は約5.7%となり、ぎりぎり有意とならない(両側検定)。実際問題として、このように正確な有意確率を求めても、例題で示したように χ^2 分布による近似を用いた場合でも、数値に大差ないことが理解されよう。したがって、比較的標本数が大きい場合、計算の簡便さからいって、McNemar 検定を使用したほうがよいだろう。