

データの順位を用いたグループの比較

高木 廣文

平成11年 8 月
超音波検査技術 Vol. 24 No. 4

統計講座 (第5回)

データの順位を用いたグループの比較

高木 廣文

はじめに

心疾患の患者と健常者の脈波速度を比較するために、超音波パルスドプラ法により上肢動脈における平均脈波速度のデータを集めたとして、このような場合、すでに説明したように、通常は二つのグループの平均値と分散（または標準偏差）を求めて、「2群の母平均値の差の検定（t検定）」を行うことが多い。しかし、忘れてはならないこととして、このt検定を用いるための前提として、「検定に用いるデータの母集団での分布は正規分布に従っている」ということが仮定されている。多くの場合、我々が得ることができるのは標本データのみであり、このため母集団でのデータの分布は分からないのが普通である。特に、極端に標本数が少ない場合には、その母集団の分布を特定することは、困難なことが多い。そのような場合、母集団での分布に、正規分布を仮定した検定方法を用いることに不安を覚えるかもしれない。正規分布を仮定しない方法があれば、そのような規制がなく、無用な心配をすることもないだろう。

母集団の分布に正規分布などの特定の分布を仮定しない方法は、「分布によらない検定」とか「ノンパラメトリック検定」などと総称されている。一方、すでに説明したt検定、F検定などは「パラメトリック検定」と呼ばれることがある。

今回は、ノンパラメトリック検定で2グループの量的データを比較する方法について説明する。

表1 健常群と動脈硬化群の平均脈波速度 (cm/sec)

健 常 群	17	20	25	26	28
動脈硬化群	11	12	15	16	19

順位に基づく検定

ここでは実際の計算手順を、例題を用いて説明しよう。表1は健常群と動脈硬化群の各5名について、平均脈波速度を測定したものである（架空例）。データは全体でも10名のものであり、母集団の分布が正規分布に従うかは定かではない。

このような場合、まず、データの大きさにより、各データに順位を付ける。順位を付ける場合、小さい順に付けるか、大きい順にするかは問題ではなく、どちらでも構わない。ただし、ここでは説明のためには、小さい順に順位を付けた場合について解説する。

表1のように2グループのデータがある場合、まったく母集団での分布（どんな分布かは特定しない）が等しいとしても、偶然の影響でどちらかが大きくなることはあり得るだろう。ここでは、そのような偶然の程度をデータの大きさの順位によって検討する。大きいデータをもつグループの順位の合計は、もう一方のグループに比べ大きくなるだろうが、それがどの程度の確率で偶然起きるかを調べれば、統計学的な検定が行えることになる。

表2 平均脈波速度の順位データ

健 常 群	5	7	8	9	10
動脈硬化群	1	2	3	4	6

表2に表1の全データを小さい順に1から順位を付けたものを示した。

健常群の順位の合計(順位和) S_0 は、

$$S_0 = 5 + 7 + 8 + 9 + 10 = 39$$

となり、動脈硬化群の順位和 S_1 は、

$$S_1 = 1 + 2 + 3 + 4 + 6 = 16$$

となる。

2グループの順位和を比べると、健常群の方が2倍以上の値となっているが、この結果だけで、平均脈波速度は健常群の方が速いとは断定できない。検定のためには、「2群の母集団での平均脈波速度は等しい」場合に、偶然の影響でデータに観測されたような偏りが生じる確率(p値)を求めなければならない。

表3 順位の組合せと順位和

順位の組合せ	順位和
1, 2, 3, 4, 5	15
1, 2, 3, 4, 6	16
1, 2, 3, 4, 7	17
1, 2, 3, 5, 6	17
⋮	⋮
⋮	⋮
⋮	⋮
4, 7, 8, 9, 10	38
5, 6, 8, 9, 10	38
5, 7, 8, 9, 10	39
6, 7, 8, 9, 10	40

実際にp値を求めるには、10個のものから5個を取り出したときに、その順位和がどのようになるかをすべて数え上げる必要がある。実際に数えると、順位和が16以下になる場合は2通りある。また、順位和が39以上になる場合も2通りあることが分かる。表3にはすべての組合せを表示している訳ではない。実際には10個のものから5個選ぶ組合せの数は、全部で ${}_{10}C_5 = 252$ 通りある。

これらのことから、一方の順位和が16以下になる下側確率 p_L は、

$$p_L = 2/252 = 0.0079$$

となる。また、一方の順位和が39以上になる上側確率 p_U は、

$$p_U = 2/252 = 0.0079$$

となる。

下側確率、上側確率の一方のみを用いた検定は片側検定となるが、通常は両者を加えた両側検定により判断するのが普通である。なお2グループの標本数が等しい場合には、下側確率と上側確率が一致するが、標本数が異なる場合には、そうとは限らない。

検定のp値としては、

$$p = 0.0079 + 0.0079 = 0.016$$

を用いればよい。この結果から、p値は5%よりも小さいので、健常群と動脈硬化群では平均脈波速度に有意差があるといえることになる。

ここで説明した方法は、「ウィルコクソン(Wilcoxon)の順位和検定」と呼ばれる方法である。場合によっては、各グループの順位和から、標本数から期待される最小値を引いた値、上記の例では、健常群の順位和16から期待される最小値(標本数 $\times$ (標本数+1) $\div$ 2)、すなわち、 $5 \times 6/2 = 15$ を引いた値、

$$U = 16 - 15 = 1$$

を求めることがある。これは、「マン-ウィットニー(Mann-Whitney)のU統計量」とも呼ばれている。

ところで、順位和やU統計量から検定のためのp値を求める計算を、その都度自分で行うのはかなり面倒である。実際には、統計学の教科書や統計数値表などには、2グループのいくつかの標本数の組について、可能な順位和やU統計量のp値(または逆にp値が5%以下になる順位和やU統計量)が掲載されており、それを利用すればよい。

しかし、標本数が多くなったり、同順位のデータがある場合の数値表は用意されていない。そのような場合には、以下に説明するように、検定のためのp値を正規近似により求めればよい。

正規近似によるp値の計算

いま、2つのグループをA、Bとし、AグループとBグループの標本数をそれぞれm、n、全体の標本数を $N = m + n$ とする。

ここでは、まずいくつかのデータの値が一致す

るようなことのない場合 (タイ tie のない場合) について説明する。

Aグループの順位和を R_A , Bグループでの順位和を R_B とする。ここで, 順位は1から N までとなる。説明を簡単にするため, $R_A > R_B$ であるとしよう。

なお, すでに説明したようにU統計量は, Aグループの標本数は m なので, 順位和 R_A から期待される最小値を引くことで, すなわち,

$$U_A = R_A - m(m+1)/2$$

により求められる。同様にBグループについては,

$$U_B = R_B - n(n+1)/2$$

と求められる。

標本数が比較的大きい場合, すべての標本数の組合せについて検定のための数値表を用意するのは大変であり, 通常は, 以下のような正規分布による近似値 z_0 を用いて検定を行う。

$$z_0 = \frac{|R_A - m(N+1)/2| - 0.5}{\sqrt{mn(N+1)/12}}$$

$$= \frac{|U_A - mn/2| - 0.5}{\sqrt{mn(N+1)/12}}$$

を求めればよい。

上記の z_0 は標準正規分布に近似的に従うものと考えられるので, その値から容易に有意確率である p 値を求めることができる。

目安としては, z_0 が1.96以上であれば, 5%水準で2グループには有意差があることになり, さらに z_0 が2.58以上であれば1%水準で有意差ありとすればよい。

なお, 上式の分子の0.5は, 離散的な順位データを連続的な正規分布へあてはめる場合, 近似がより良くなるように行うもので, 「連続性の修正(補正)」などと呼ばれている。

正規分布を用いた p 値の計算の近似は, 2グループの標本のうち小さい方が4程度でもかなりよく, 8以上ならば実用上十分であるといわれている。

次に, データに同順位のある場合 (タイのある場合) について, 若干の説明を加えよう。

一般に, 順位と検定もしくはU統計量による検定のための数値表は, タイのない場合について作成されている。したがって, タイのある場合にはそれを使用することはできないので, 自分で順位和の分布をすべて数え上げ, p 値を求めるか, 以下のように標準正規分布による近似を用いる必要がある。

全体のデータのうち, 異なるデータの種類の K 個あるものとし, 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。第 i 番目に大きな値をとるデータには, すべてその順位として s_i を与えるものとしよう。すなわち,

$$s_i = t_1 + \dots + t_{i-1} + (1 + t_i)/2$$

のように順位を付ける。

次に, Aグループの順位和 R_A を計算し, 検定のための統計量 z_0 ,

$$z_0 = \frac{|R_A - m(N+1)/2| \times \sqrt{12N(N-1)}}{\sqrt{mn \times \{N^3 - N - \sum_i (t_i^3 - t_i)\}}}$$

を求める。ここで, $\sum_i$ はすべてのタイについて, 和の計算を行うことを示す。

タイのない場合には, $t_1 = t_2 = \dots = t_k = 1$ とおくと, 前述したタイのない場合の式と一致する。なお, U統計量を用いる場合でも, 同様に検定が行える。検定の方法はタイのない場合と同様の手順であり, p 値が0.05以下であれば, 5%水準で有意差があることになる。なお, タイのある場合には連続性の補正は行えない。また, 上式による正規近似は, 同順位に含まれるケース数が多すぎると精度が悪くなることが知られており, 特定のタイのデータの個数が極端に多い場合, 正規分布による近似を用いるべきではない。

表1のデータについて, 正規近似により順位和検定を行ってみよう。すでに求めたように, 健常群の順位和は39, 標本数はそれぞれ5例であり, データにタイはない。したがって,

$$z_0 = \frac{|39 - 5 \times (10 + 1) / 2| - 0.5}{\sqrt{5 \times 5 \times (10 + 1) / 12}}$$

$$= 2.298$$

となる。この値は1.96より大きく、2.58よりは小さいので、2グループ間の平均脈波速度には5%水準で有意差があると検定できる。実際には、標準正規分布で2.298の両側確率を求めると、p値は約0.022となり、正確な値である0.016とほとんど差がないことがわかる。

おわりに

最後に、表1のデータを通常の2群の母平均値の差の検定(t検定)をしてみるとどうなるかについて、実際にやってみよう。計算の手順は省略するが、検定のためのt値は3.453となり、自由度8のt分布表から、このp値は約0.009と求められる。順位和検定よりも有意確率は小さくなっていることが分かるだろう。このように、母集団分布に正規分布を仮定するt検定は、仮定が正しいとすれ

ば、一般的に統計学的検定ではより有意性の高い結果を導く傾向がある(検定の効率が高い)。ただし、母集団の分布に特定の仮定を置かない検定であるノンパラメトリック法の適用範囲は、正規性の仮定がいないため、より広いといえよう。また、検定結果も順位のみをデータとしている割には、それほど効率の悪い方法ではない。正規性の仮定に疑問がある場合には、順位和検定を用いるべきであろう。

今回の方法も正規近似によるp値ではあるが、著者のホームページ(<http://www.clg.niigata-u.ac.jp/~takagi/>)で簡単に計算できるので実際に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) 豊川裕之，柳井晴夫編著：医学・保健学の例題による統計学，現代数学社（京都）1982
- 3) E. L. レーマン原著，鍋谷清治他訳：ノンパラメトリックス，森北出版（東京）1978（E. L. Lehmann：Nonparametrics, Holden-Day, 1975）
- 4) 高木廣文：2グループの平均値を比べる，超音波検査技術 1998；23：400-405