統計入門　〔第8回〕

棄却検定・分布によらない検定

高木廣文*　新田裕史**　中井里史***

I. スミルノフーグラブス検定

データを収集した場合，その中に他のものと極端に数値の異なるものが含まれることがある。そのような，小さすぎる値か大きすぎる値をもつデータは“**外れ値(outlier)**”と呼ばれる。外れ値は偶然の変動によって起こることもあり，たんに他のデータと数値が離れているからといって，母集団に仮定される正規分布から逸脱した“異常値”として，分析から除外することはできない。

いま，$x_1, x_2, \cdots\cdots, x_n$, と$n$個のデータがあり，その平均値を$\bar{x}$，不偏分散を$s^2$としよう。外れ値を異常値として棄却するためには，

$$t_i = |x_i - \bar{x}| / s$$

を計算し，$max\,(t_i)$を求めればよい。この値と検定のための数値表[1]の$\alpha/2$％点の値を比較し，t_iが大きければ有意水準α％で棄却する。したがって，5％水準で検定したい場合には，数値表の2.5％点と比較することになる。これが，**スミルノフーグラブス検定**である。

いま，10人の血液のpHを測定したところ，

7.30　7.37　7.39　7.40　7.41　7.42　7.42
7.43　7.44　7.44

となっていたとする。平均値は7.402，不偏分散は0.0017733（標準偏差は0.04211）である。$t_1 = |7.30 - 7.402| / 0.04211 = 2.422$となり，この値が10個のデータのうち最大のものである。棄却検定の数値表で標本数10，有意水準2.5％の値を求めると2.290である。したがって，7.30というデータは，他のデータと同一正規分布からの値とはみなせず，5％の危険率で異常値と考えられ棄却することができる。

表1　A薬とB薬の症状改善について

	著効	有効	不変	悪化	計
A薬	8	10	6	1	25
B薬	3	9	10	4	26
計	11	19	16	5	51

II. 順位和検定とU検定

表1は，2種の薬剤A，Bをある疾患の患者に無作為に投与し，その効果を調べたものである。このデータから，二つの薬剤の効果に差があるかを，どのように検定すればよいだろうか。まず，簡単にその方法を説明する。

A群とB群の標本数をm，nとし，$N = m + n$とする。まず，検定しようとする変数に関して，データを大きさの順番にならべかえる。この場合，順番は群ごとではなく全体を大きさの順に並べ，1からNまで順位をつける。まず同順位がない(“タイ tie”のない場合）について説明する。

A群でのデータの順位を$s_1, s_2, \cdots\cdots, s_m$とし，B群での順位を$r_1, r_2, \cdots\cdots, r_n$とする。これらの順位の合計，すなわち“順位和”をA群ではR_A，B群はR_Bとする。順位は1からNまでなので，二つのグループの順位和を足したものは，$N(N+1)/2$となる。説明を簡単にするため，$R_A > R_B$

* 文部省統計数理研究所
〔〒106 東京都港区南麻布 4-6-7〕
** 国立環境研究所
*** 東京大学医学部保健学科疫学教室

とする。

2群の分布の形状は同一であるが，その位置に差があるかを検定するためには（一方がより高い方に分布がずれているか否かは），適当な定数 c_1, c_2 を用いて，

$R_A \geqq c_1$ または $R_B \leqq c_2$

の場合，有意差ありと検定できる。これが，“**ウィルコクソン（Wilcoxon）の順位和検定**”と呼ばれる方法である。c_1 と c_2 の値については，有意水準を5％や1％などとした場合で，2群の標本数が20以下ぐらいまでについて，その値を示す数値表が用意されている[1]。また，場合によっては，以下のように順位和を変換して用いることもある。A群の標本数は m なので，順位和 R_A から期待される最小値を引くと，

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

となる。同様に，B群については，

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

となる。U_A と U_B は“**マン-ウィットニー（Mann-Whitney）のU統計量**”と呼ばれるものである。

標本数が比較的大きい場合，正規分布による近似を用いて，

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

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

を求める。z_0 は標準正規分布に従うので，z_0 と両側5％点1.96か1％点2.58を比較すればよい。z_0 の方が大きければ，有意差があるとする。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，“連続性の補正”と呼ばれる。

次に，タイのある（同順位のある）場合について説明する。一般に，順位和検定もしくは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。タイのある場合，順位和の分布を全て数えあげ，確率計算をするか，以下のように標準正規分布による近似を用いることになる。

全体のデータのうち，異なる値は k 種類あるものとしよう。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，

$$s_i = (t_i+1)/2 + \sum_{j=0}^{i-1} t_j \quad (i=1, 2, \cdots\cdots, k)$$

のように順位をつける（ただし，$t_0=0$）。

次に，A群の順位和 R_A を計算し，検定のための統計量，

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

を求める。タイがない場合，$t_1 = t_2 = \cdots\cdots = t_k = 1$ となり，上式はタイのない場合の式と一致する（連続性の補正のない場合）。なお，U 統計量を用いる場合も同様である。検定の方法はタイのない場合と同様に，z_0 が1.96より大きければ，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わない。なお，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。したがって，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のようなクロス表は，一般に χ^2 検定がよく用いられる。しかし，効果を表す四つのカテゴリに順序がつけられると考えられるので，タイのある場合の Wilcoxon の順位和検定を行う。“著効”には順位として6を，“有効”には21，“不変”には38.5，“悪化”には49を与える。

A群の順位和は，

$R_A = 8 \times 6 + 10 \times 21 + 6 \times 38.5 + 1 \times 49 = 538$

となり，その期待値は，$E(R_A) = 25 \times (51+1)/2 = 650$ となる。

また，分散は，

$$V(RA) = \frac{25 \times 26}{12 \times 51 \times 50} \times [(51^3 - 51) - \{(11^3 - 11) + (19^3 - 19) + (16^3 - 16) + (5^3 - 5)\}] = 2554.118$$

となる。これらの値から，

$$z_0 = \frac{|538 - 650|}{\sqrt{2554.118}} = 2.216$$

と求められる。z_0は1.96より大きいので，5％水準で有意差が認められる。この結果，A薬はB薬に比べ，症状改善に有効なものと考えられる。なお，クロス表の検定としてχ^2検定を用いると，χ^2値は5.108となり，その有意確率は0.164（16.4％）となり，有意とはならない。順位和検定（またはU検定）では，カテゴリ間の順序に関する情報を用いるため，効率のよい方法である。

III. 一般化ウィルコクソン検定

ある疾患の患者を無作為に2群に分け，一方に治療法Aを他方にBを行い，経過を観察した。5年間の経過観察の結果は，**表2**に示したようになった。この結果から二つの治療法による生存率に差があるといえるだろうか。なお，表中の打ち切り例とは，死亡が確認される以前に調査期間が終了した場合と，追跡が不可能になった場合を含んでいる。

このような研究では，患者は調査の開始から全員が参加するわけではなく，随時対象として追加されるので，実際の観察期間には患者により開きがある。分析のためには，観察開始時期を同一時点に揃え直し，以後の生存数，死亡数，また観察打ち切り例数などを整理する必要がある。例えば，観察が1年間のみの患者の場合，それ以後の期間はデータがないので，観察期間以降については打ち切り例として処理されることになる。

2群の生存率の差を分析するためには，各種の方法が考案されているが，ここでは**“一般化ウィルコクソン検定”**を説明する。

A群の標本数をmとし，各症例の生存期間（死亡確認までの期間）を$x_1, x_2, \cdots\cdots, x_k, x^*_{k+1}, \cdots\cdots, x^*_m$とする。ここで，$x^*$は打ち切りもしくは脱落のあった例を示す。同様に，B群の標本数をnとし，各症例の生存期間を$y_1, y_2, \cdots\cdots, y_l, y^*_{l+1}, \cdots\cdots y^*_n$とする。

B群の各症例y_jまたはy^*_jに対する，A群の各症例x_iまたはx^*_iの生存期間の長短により，得点u_{ij}を以下のように定める。

$$u_{ij}=\begin{cases}1：x_i>y_j，\text{または } x^*_i\geqq y_j\\ -1：x_i<y_j，\text{または } x_i\leqq y^*_j\\ 0：\text{その他の場合}\end{cases}$$

表2 2群の生存期間

確認例（死亡例）				打ち切り例*			
期間	A群	B群	計	期間	A群	B群	計
1	16	12	28	1+	17	13	30
2	11	12	23	2+	16	16	32
3	9	12	21	3+	5	9	14
4	6	5	11	4+	5	8	13
5	1	1	2	5+	8	10	18
計	43	42	85	計	51	56	107

* 打ち切りは期間の間に起こっているので各期間の数値に“+”をつけた。

A群の得点の総計をWとしよう。

$$W=\sum_{i=1}^{m}\sum_{i=1}^{n}u_{ij}$$

である。

ところで，2群の各症例を比較して得点を付けるかわりに，A群とB群を一緒にして，$N(=m+n)$個の対象について考える。

各症例z_iもしくはz^*_iに比べて，

$\begin{cases}s_i=\text{生存期間の短い例数}\ (u_{ij}=1\text{に対応})\\ r_i=\text{生存期間の長い例数}\ (u_{ij}=-1\text{に対応})\end{cases}$

をそれぞれ求め，これらの差，$w_i=s_i-r_i$を考える。

A群の症例についてw_iを求めて合計すると，これはWと一致する。Wの分散$V(W)$は，上記のw_iをA群とB群の全てについて求め，

$$V(W)=\frac{mn}{N(N-1)}\sum_{i=1}^{N}w_i^2$$

により求められる。

統計量Wとその分散$V(W)$を用いて，

$$z_0=\frac{|W|-1}{\sqrt{V(W)}}$$

を求め，z_0が標準正規分布に従うことにより，2群の生存期間の差を検定する。すなわち，z_0が1.96より大きければ，5％水準で有意差ありとすればよい。また，2.58よりも大きければ1％水準で有意差があることになる。これが，一般化ウィルコクソン検定と呼ばれる方法である。

表2をもとに，各期間に含まれる症例に対して与える得点を求める。“期間1”の死亡例に対しては，これより生存期間の短い例はないので，s_1

表 3　各症例に与える得点とその2乗

確認例			打ち切り例		
期間	得点	得点^2	期間	得点	得点^2
1	−164	26,896	1+	28	784
2	−83	6,889	2+	51	2,601
3	−7	49	3+	72	5,184
4	39	1,521	4+	83	6,889
5	65	4,225	5+	85	7,225

=0となり，また，期間の長い例数は，$r_1=(85-28)+107=164$となる。得点w_1は，s_1とr_1の差であるから，$w_1=0-164=-164$となる。“期間2”については，$w_2=28-(21+11+2+32+14+13+18)=-83$となる。同様にして他の期間についても求められる。

打ち切り例については，打ち切り例同士での比較は行わないので，“期間1＋”は，$w_{1+}=28-0=28$となり，また“期間2＋”は，$w_{2+}=28+23-0=51$と求められる。他も同様であり，このようにして求めた得点とその2乗の値を**表3**にまとめた。

表3をもとに，A群の症例の得点の合計Wを求める。

$$W=16\times(-164)+11\times(-83)+9\times(-7)+6\times39+1\times65+17\times28+16\times51+5\times72+5\times83+8\times85=-554$$

となる。Wが負なので，A群の方が生存期間は短いようである。次に，Wの分散$V(W)$を求めるが，そのために各症例の得点の2乗和を求めておく。

$$2乗和=28\times26896+23\times6889+21\times49+11\times1521+2\times4225+30\times784+32\times2601+14\times5184+13\times6889+18\times7225=1336680$$

となる。A群は94例，B群は98例，総計192例であり，分散$V(W)$は，

$$V(W)=\frac{94\times98}{192\times(192-1)}\times1336680=335773.7827$$

となる。これらの値からz_0は，

$$z_0=\frac{|-554|-1}{\sqrt{335773.7827}}=0.954$$

となる。z_0は1.96と比べて小さいので，5％水準で2群の生存期間に差があるとはいえないことになる。

今回取り上げた内容について，詳細な説明を知りたい読者のために，また検定に必要な数値表について，最後に文献を示しておいたので，参照して欲しい。

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) 竹内　啓，大橋靖雄：統計的推測—2標本問題．日本評論社，1981.
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，1978 (Lehmann E.L.: Nonparametrics, Holden-Day, 1975).
4) 富永祐民：治療効果判定のための実用統計学—生命表法の解説—，蟹書房，1980.

*　　*　　*