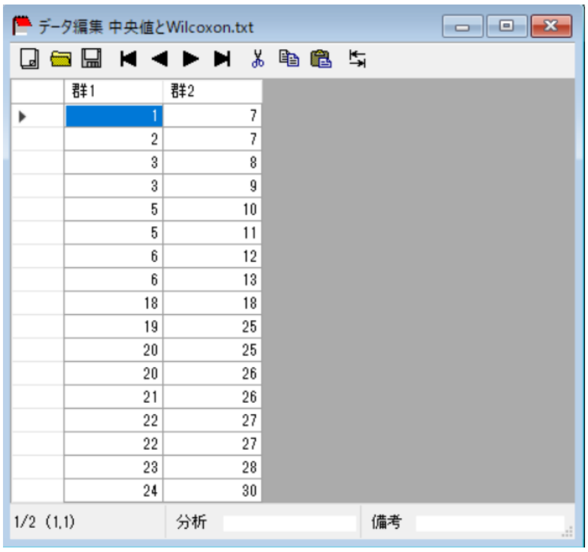
中央値と Wilcoxon 順位和検定について

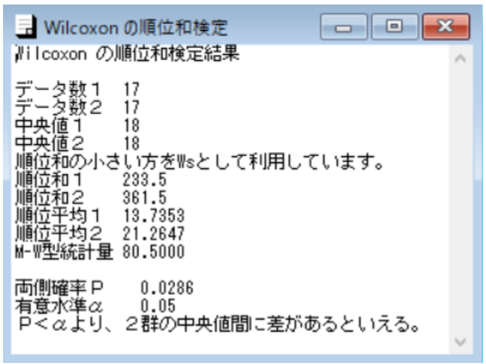
Wilcoxon 順位和検定は中央値の比較検定と解釈されることが多いが、実際のところ何を比較しているのだろうか。データを元に調べてみる。

この問題で私が最初に不思議に感じたのは以下の事例です。あるデータとその基本統計量を求めた結果を図 1 に示します。


データ編集 中央値とWilcoxon.txt

群1	群2
1	7
2	7
3	8
3	9
5	10
5	11
6	12
6	13
18	18
19	25
20	25
20	26
21	26
22	27
22	27
23	28
24	30

1/2 (1,1) 分析 備考


Wilcoxon の順位和検定

Wilcoxon の順位和検定結果

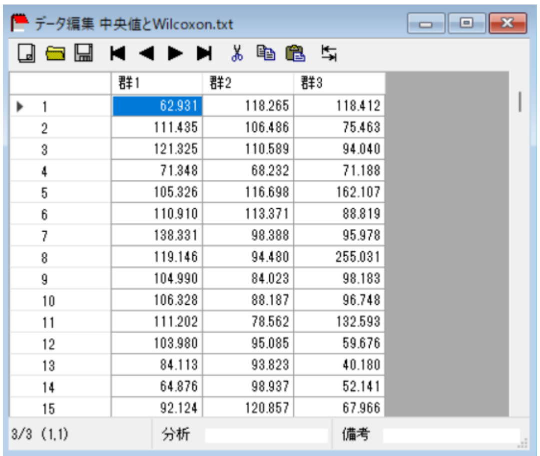
データ数１ 17
データ数２ 17
中央値１ 18
中央値２ 18
順位和の小さい方をWsとして利用しています。
順位和１ 233.5
順位和２ 361.5
順位平均１ 13.7353
順位平均２ 21.2647
M-W型統計量 80.5000

両側確率Ｐ 0.0286
有意水準α 0.05
Ｐ＜αより、２群の中央値間に差があるといえる。

図 1　データと中央値

この例では、2 群のデータの中央値が共に 18 なのに、Wilcoxon 順位和検定の結果に差の有意性が示されています。t 検定に慣れていると、t 統計量の中で平均の差を取りますので、このようなことは絶対にありません。しかし、順位和を使った検定ではこのようなことが起こります。これは順位和検定が中央値だけでなく分布の形状を見ている結果のように思えます。果たして一般に順位和検定は中央値の検定と言ってもよいのでしょうか。

これを調べるために、図 2 のように 3 つのデータセットを用意します。右に基本統計量を表しています。


データ編集 中央値とWilcoxon.txt

	群1	群2	群3
1	62.931	118.265	118.412
2	111.435	106.486	75.463
3	121.325	110.589	94.040
4	71.348	68.232	71.188
5	105.326	116.698	162.107
6	110.910	113.371	88.819
7	138.331	98.388	95.978
8	119.146	94.480	255.031
9	104.990	84.023	98.183
10	106.328	88.187	96.748
11	111.202	78.562	132.593
12	103.980	95.085	59.676
13	84.113	93.823	40.180
14	64.876	98.937	52.141
15	92.124	120.857	67.966

3/3 (1,1) 分析 備考


基本統計量

	群1	群2	群3
データ数	10000	10000	10000
最小値	18.659	64.604	20.31
最大値	165.734	124.596	352.22
平均値	94.265	94.425	101.337
中央値	94.276	94.3305	94.3675
レンジ	147.075	59.992	331.91
分散	392.765	306.207	1587.533
標準偏差	19.818	17.499	39.844
不偏分散	392.804	306.237	1587.692
標準偏差	19.819	17.500	39.846
標準誤差	0.198	0.175	0.398
第1四分位数	80.790	79.087	72.405
第3四分位数	107.752	109.825	122.656
四分位範囲	26.962	30.738	50.251
歪度	-0.001	0.020	1.128
尖度	-0.075	-1.222	2.013

図 2　3 つのデータセット

データ数はいずれも 10,000 個で、中央値がほぼ同じです。しかし、平均値や標準偏差など

は大きく違っています。３つのデータをヒストグラムで表すと、図 3 のようになります。

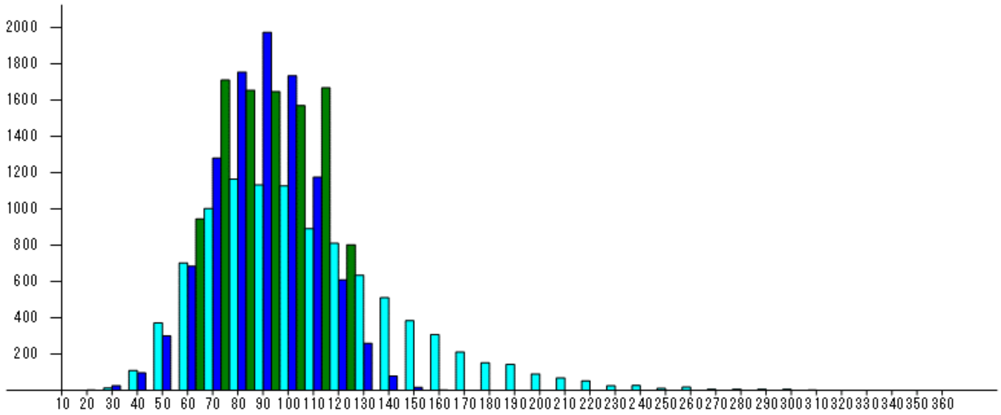


図 3　3 群をまとめたヒストグラム

分布の形状としては、群 1 が正規分布、群 2 が一様分布、群 3 が対数正規分布です。

Wilcoxon の符号付き順位和検定

まず、指定値を 94.3（ほぼ中央値と同じ）として、Wilcoxon の符号付き順位和検定を行ってみましょう。

群 1		群 2		群 3	
データ数	10000	データ数	10000	データ数	10000
指定値	94.3	指定値	94.3	指定値	94.3
中央値	94.2760	中央値	94.3305	中央値	94.3675
中央値－指定値	-0.0240	中央値－指定値	0.0305	中央値－指定値	0.0675
利用数	10000	利用数	10000	利用数	10000
最小順位和	24964391.5	最小順位和	24789328.5	最小順位和	22458551.5
両側確率Ｐ	0.8950	両側確率Ｐ	0.4603	両側確率Ｐ	0.0000

群 1 と群 2 については、中央値の検定と呼んでも良さそうですが、群 3 については中央値だけを比較していると考えることは困難です。このように、この検定では、分布が中央値に対して左右対称なら中央値の検定と考えることができますが、分布が非対称の場合、中央値の検定と考えるには無理があるようです。

Wilcoxon の順位和検定

次にこれを 2 群ごとに Wilcoxon 順位和検定を用いて比較してみましょう。

表 1　Wilcoxon の順位和検定結果

群 1 と群 2		群 2 と群 3		群 3 と群 1	
データ数１	10000	データ数１	10000	データ数１	10000
データ数２	10000	データ数２	10000	データ数２	10000
中央値１	94.276	中央値１	94.3305	中央値１	94.3675
中央値２	94.3305	中央値２	94.3675	中央値２	94.276
順位和１	99828178	順位和１	98657374	順位和１	101597094.5
順位和２	100181822	順位和２	101352626	順位和２	98412905.5

順位平均１　9982.8178 順位平均２　10018.1822 M-W 型統計　49823178.0 両側確率Ｐ　　0.6649	順位平均１　9865.7374 順位平均２　10135.2626 M-W 型統計量　48652374.0 両側確率Ｐ　　0.0010	順位平均１　10159.7095 順位平均２　9841.2906 M-W 型統計量　51592094.5 両側確率Ｐ　　0.0001

この結果と補遺の順位和の取り方を考えますと、分布が左右対称である者同士、または分布の型が同じ場合、Wilcoxon の順位和検定は中央値の検定と考えてもよく、分布の型が異なる場合は、中央値の検定と考えるには無理がありそうです。

経済の分野では主にパラメトリック検定が使われ、医療系ではノンパラメトリック検定も使われているという現実がありますが、これは単に経済分野ではデータ数が多く取れるから中心極限定理によりパラメトリック検定が利用可能というだけではありません。経済分野では２つの群の分布の型が同じだろうと想定しているように思えます。同じ分布の型なら、パラメトリックな検定とノンパラメトリックな検定は、比較するものが平均値か中央値の違いですので、どちらでも大差ないように思えます。しかし、医療分野では、２つの分布の型が違うことも大いに想定されます。この部分を調べることも大切なので、医療分野でノンパラメトリック検定が使われるのではないでしょうか。

補遺

Wilcoxon の符号付順位和検定について

Wilcoxon の符号付き順位和検定について例を用いて簡単に説明しておきます。今、図 A1 のような 3 種類のデータを考えます。上のデータ並びをケース A、中央のデータ並びをケース B、下のデータ並びをケース C と呼び、状況の違う 3 つのケースとします。この 3 つのケースのうち、代表値（平均値や中央値）が μ から最もずれているのはどれでしょうか。

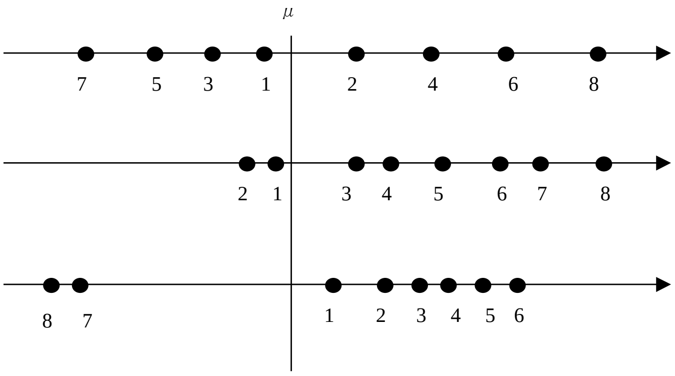


図 A1　検定概念図

一番上のデータは平均から左右均等に散らばっていますので、これは違います。しかし真ん中のデータは極端に右に偏っていますので、これだと分かります。視覚的にはすぐに分かりますが、数値的には何を使ってずれを判定するのでしょうか。一番下のデータはあまりずれていないように感じますが、平均から右にずれている個数は 2 番目と同じなので、左右の個数ではありません。

このデータに対して左右に関係なく 0 に近いところから順番に番号（順位）を付けてやる

ことにします。それが上の図に付いた番号です。この番号を 0 以上と 0 未満のところで合計します。それを表にすると表 A1 のようになります。

表 A1　符号付き順位和

	－群	＋群
ケース A	16	20
ケース B	3	33
ケース C	15	21

真ん中のデータは合計が極端に違います。この番号（μ に近い順位）の和によってデータの偏りをみる検定が Wilcoxon の符号付き順位和検定です。実際に利用する式は以下です。この検定は中央値の検定と言われますが、分布が左右非対称の場合は、結果は分布の型にも依存します。

理論

指定値の左右の順位和を求め、その小さい方を R とする。

標本数が多いとき

$$z = \frac{|R - n(n+1)/4| - 1/2}{\sqrt{n(n+1)(2n+1)/24}} \sim N(0,1)$$ 分布（正の部分）　（Yates の連続補正）

　ソフトではさらに、同順位の補正というものも入る。

標本が少ないとき

　数表を利用する。

Wilcoxon の順位和検定について

　Wilcoxon の順位和検定について例を用いて簡単に説明しておきます。今データを白丸で表した 1 群と黒丸で表した 2 群の 2 つの群を考えます。下図の上のデータ並びをケース A と下のデータ並びをケース B と呼び、状況の違う 2 つのケースとします。さてどちらのケースの代表値（平均値や中央値）が異なっているように感じるでしょうか。

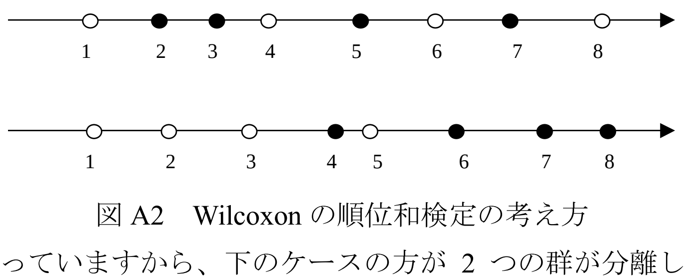


図 A2　Wilcoxon の順位和検定の考え方

上は白と黒が混ざっていますから、下のケースの方が 2 つの群が分離しているように感じます。この感覚をどのように表現するのでしょうか。

　今左から順番に番号を付けて行き、白丸と黒丸とでその番号の合計を取り、表 A2 を作ります。

表 A2　順位和

	白丸群	黒丸群
ケース A	19	17
ケース B	11	25

この表を見ると、ケース A では白丸群と黒丸群はほぼ同じ順位和ですが、ケース B では 2 つの群で順位和はかなり違います。この違いを利用して 2 群の中央値を比較する検定を Wilcoxon の順位和検定といいます。具体的には以下の通りです。この検定も中央値の検定と言われますが、2 つの群で分布の型が違う場合は、結果は分布の型にも依存します。

理論

両群のデータの小さい順に順位を付け、データ数の少ない群（$n_1 \le n_2$）の順位和をWとする。但し、同じ値のデータにはそれらが異なると考えた場合の順位の平均値を付ける。例えば同順位の 3 位には(3+4)/2=3.5 の順位を付ける。

データ数が多い場合　両群の中央値が等しいとすると

$$z = \frac{|W - n_1(n_1 + n_2 + 1)/2| - 1/2}{\sqrt{n_1 n_2 (n_1 + n_2 + 1)/12}} \sim N(0,1)$$ 分布（正の部分）　（Yates の連続補正）

データ数が少ない場合

数表を利用する。