

ノンパラメトリック検定法

60151039 東 裕之輔

資料

▶ 医薬研究者の視点からみた道具としての統計学 第12章 「ノンパラメトリック検定法」

奥田千恵子 著

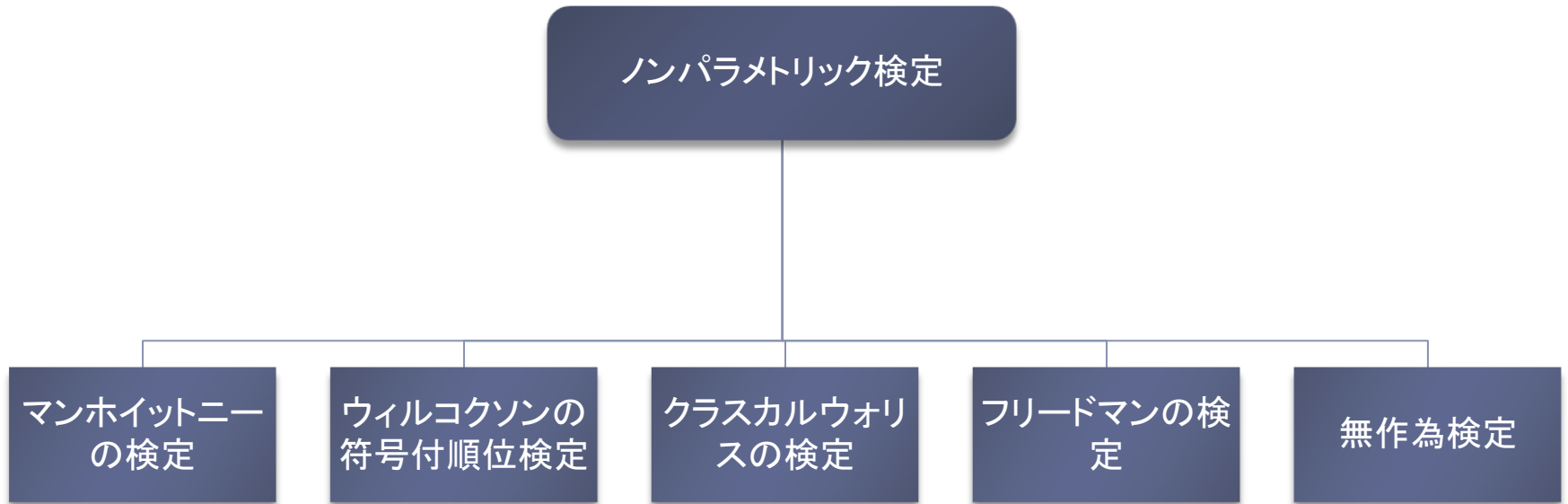


ノンパラメトリック検定法

- ▶ 母数(parameter)を用いない検定法
- ▶ 2グループの比較をデータの順位に基づいて行う方法
 - ▶ ノンパラメトリック検定ではデータに順位をつける。順位和とはその順位の合計。
- ▶ 標本数が少なく、正規性が確認できないときに使用する
- ▶ 測定値の数値自体に信頼性が低い



ノンパラメトリック検定法の種類



マン・ホイットニーの検定

▶ 2群間の対応のないデータの順位和を比較する方法。

▶ この検定を行う条件

1)標本は母集団から無作為抽出されている。

2)標本が互いに独立している。

▶ 1)は統計解析法全てに求められているもの。

▶ 2)が満たされていない場合は、ウィルコクソンの符号付順位検定へ

▶ 統計ソフトではマン・ホイットニーのU検定を用いる。



例題

A,B2種類の治療を行い、一定の基準で患者の症状をスコアで表したデータが以上になるとすると両群の順位和は？

両群のデータを混ぜて、小さいほうから順位をつけると、

治療A		治療B	
データ	順位	データ	順位
6	1	19	3
11	2	27	5
25	4	33	6

※同順位であれば5位と6位が同じ値ならば5.5位とする。

A群の順位和： $T_a = 1 + 2 + 4 = 7$

B群の順位和： $T_b = 3 + 5 + 6 = 14$

順位和の確率

全ての組み合わせの順位和を求める

	Aの順位	A群の順位和
1	1,2,3	6
2	1,2,4	7
3	1,2,5	8
4	1,3,4	8
5	1,2,6	9
6	1,3,5	9
7	2,3,4	10

- ・起こりうる順位の組み合わせを全て洗い出すと $6C3=20$ 通り
- ・そのうち7以下なのは2つだけなので、確率は $2/20=0.1$
- ・0.1という値はP値となる。

P値

- ▶ P値とは偶然生じる確率。
- ▶ 今回、 $2/20=0.1$ が片側検定でのP値となり、両側検定では0.2である。
- ▶ 有意確率は5%以下なので今回は統計的有意差は認められないという結果になる。



ウィルコクソン符号付き順位検定

- ▶ 対応のあるt検定のノンパラメトリック検定法
- ▶ 個々のデータの差の代わりに、対応のあるデータの差を用いる。
- ▶ 順位和の分布を正規分布に近似して求める。
- ▶ 条件
 - 1) 標本が何らかの形で対になっている。
 - 2) 対になった標本は母集団から無作為抽出されている。



例題

8人の患者の治療の前後の症状をスコアで表したデータが以下のようになった。治療の前後で差はあるか？

患者	治療前	治療後	差	符号付き差の順位
1	23	36	13	4
2	18	19	1	1
3	55	53	-2	-2
4	78	89	11	3
5	21	21	0	
6	49	64	15	5
7	39	67	28	7
8	44	65	21	6

符号付順位・・・差が0に近い順に順位づけする。差が負の場合は負の数で付ける。
差が0のものは順位づけしない

順位和を求める

- ▶ 求める順位和
 - ▶ プラス値だけの $T+$
 - ▶ マイナス値だけの $T-$
 - ▶ 差が0以外のペアの数 m
- ▶ $T+ = 4 + 1 + 3 + 5 + 7 + 6 = 26$
- ▶ $T- = -2$
- ▶ $m = 7$



Z値計算

- ▶ Z値の計算にはT+,T-のどちらか小さい方Tと、mを以下の式に代入し、求める。

$$Z = \frac{T - \frac{m(m+1)}{4}}{\sqrt{\frac{m(m+1)(2m+1)}{24}}}$$

- ▶ 後は標準正規分布からP値が決まる。
 - ▶ 例題の場合Z=-2.028と求まるので、そのときのP値は0.043
 - ▶ 両側検定で有意水準が5%の場合、治療前後のスコアに差があるといえる。



クラスカルウォリスの検定

▶ ノンパラメトリック分散分析法

- ▶ 3群間以上で比較するときに使用する。
- ▶ 3つ以上の群間で順位付けし、各順位和を求めて比較する。

▶ 手順

1. 両群のデータを混ぜて順位を付ける。
2. データの小さいものから昇順に順位を付ける。
3. 同じ値であれば平均の順位(5位=6位ならば共に5.5位とする)
4. 各群の総和を求める。
5. H_2 統計量を求める。(式はマンホイットニーのU検定と同じ)
6. X分布からP値を求めて仮説を検定する。



例題

A,B,Cの治療を受けた患者のデータが以下ようになった。

A		B		C	
データ	順位	データ	順位	データ	順位
45	14	36	9	29	6
39	11	44	13	12	1
40	12	19	3	24	5
51	15	23	4	37	10
35	8	34	7	16	2
TA	60	TB	36	TC	24

検定統計量Hを求める $H=6.270$ $P=0.035$ よって有意差あり。
HはマンホイットニーのU検定の時のU値のこと。

フリードマンの検定

▶ ノンパラメトリックでの反復側転分散分析

- ▶ 同じ対象から経時的に、あるいは同じ対象に異なった処置をして、かつ3つ以上標本がある時に使用。
- ▶ 例：時間経過で違いはあるか

▶ 手順

1. 個体ごとに他の個体とは無関係に、各時点で順位付け
2. 各時点毎に全個体の順位和を求める。
3. 順位和の分布を表す検定統計量 X^2 値を求める。
4. X^2 値から X^2 分布からP値を求め仮説を検定する。



例題

被験者	治療前		1週間後		2週間後	
	データ	順位	データ	順位	データ	順位
1	23	1	36	2	48	3
2	18	1	19	2	21	3
3	55	3	43	1	49	2
4	78	1	89	2	93	3
5	21	1.5	21	1.5	25	3
6	49	1	64	2	73	3
7	39	1	67	2	89	3
順位和		9.5		12.5		20

$$X=8.375$$

$$P=0.015$$

よって有意水準が5%の場合、 $P \leq 0.05$ となり有意差が

▶ 求められる。

無作為検定

- ▶ 今までのものは母集団から無作為抽出されていることを前提であった。
- ▶ 母集団という概念を一切用いないのが無作為検定。
- ▶ 手順
 1. 両群のデータを混ぜ、再び2群に分けるすべての組み合わせを列挙する。(とても計算量多い)
 2. それぞれの組み合わせで平均値を求めて差を計算する。
 3. 有意水準、片側検定か両側検定を決める。
 4. 現実の平均値の差以上の差が出る組み合わせの出現確率、Pを計算し、仮説の検定を行う。



例題

治療A	治療B
6	19
11	27
25	33

左の数値例を無作為検定を用いて比較せよ

↓ 全ての組み合わせ20通りについて算出していく

	群	データ	平均値	平均値の差(B-A)
1	A	6 11 19	12	16.333
	B	25 27 33	28.333	
2 実際のデータ	A	6 11 25	14	12.333
	B	19 27 33	26.333	
3	A	6 11 27	14.667	11
	B	19 25 33	25.667	
4	A	6 19 25	16.667	7
	B	11 27 33	23.667	

解説

- ▶ 全ての組み合わせで20通りの中で実際のデータより平均値の差が大きいものは1番目と2番目しかない。
- ▶ つまり平均値の差が実際の値より大きくなる確率は $2/20=0.1$
- ▶ これをP値とすると有意水準が5%の場合、有意差はみとめられない。



無作為検定の現状と今後

▶ 現状

- ▶ あまり母集団を想定した検定法の結果と変わりはない。
- ▶ 母集団を想定することの矛盾が放置されている。
- ▶ 無作為検定は現時点では必要ないかもしれない。

▶ 今後

- ▶ 無作為検定ができるソフトもが普及し始めている
- ▶ 無作為検定の値を近似値と考えていくのが、大方専門家の意見である。

