

ケンドールの一致係数Wの算出

Die Fuehrung von Kendall's Konkordanzkoeffizient

中野嘉弘 (札幌市南区、88歳)

NAKANO Yoshihiro (Sapporo, JAPAN)

yoshihiro@river.ocn.ne.jp FAX 専 011-588-3354

鈴木義一郎先生の旧著「データ解析術」を、最近、復習した話である。

0. は し が き

Yahoo 知恵袋なる面白い Web 番組がある。歴史部門や数学・物理部門が中々、有益である。十日ほど前 (Mar/12 14:38) に、「ケンドールの一致係数Wの算出式を教えてください!!」と次の如き質問が出現した。

「3人の評定結果の一致率を計算したいのですが、統計が苦手なので、単純にHPを見ても、計算式の記号 (RとかTとか)の意味がわからず、式が理解できません。

一致係数Wを算出するために必要なデータ情報と、計算式のアルファベットの意味 (Rは何を表していて、Tは・・・)を教えてくださいましたら幸いです。

よろしく申し上げます!!

1. 原 質 問 への回答

質問日時: 2010/3/12 14:38:53
残り時間: 5日間
投票開始: 2010/3/20 02:38:58
投票数: 1
回答数: 1
お礼: 知恵コイン 50枚
閲覧数: 16

回答 (1件中1?1件) nakanochurchさん

統計学では、勝手な記号を使うのが常らしいので、困りますね。
私の回答が、何処まで、お役に立つか? 心配です。

話は順位相関係数ですね。

R と云えば、Spearman's rho ρ 。 T と云えば、Kendall's tau τ 。

ケンドールの一致係数 Kendall's coefficient of Concordance は W。
W=1 完全一致、W=0 完全不一致。

テキストによって、記号が異なるのが常ですので、うっかり、答えられない。

質問者のお持ちの資料の定義式らしきものの片鱗をお示し下さい。
その後、或いは、話は進展するかも？

統計の参考書は、なるべく、平易なものを、ご覧なさいませ。
鈴木義一郎著「データ解析術」実教出版、1989 第12版 pp.96-103 が最も
判り易い。 ISBN4-407-02181-0 C1034 \1200

東南アジアで好評のカナダ産の J 言語グループ (日本) では、私の研究仲間
(文部省・統計数理研究所) の ベテラン教授でした。

編集日時：2010/3/16 14:48:03 回答日時：2010/3/16 14:12:31

実は、今回の推薦書「データ解析術」は、大昔の平成元年、私がたまたま、
講義に採用した、好著であったのだ。鈴木先生とは、古いおなじみである。

2. 順位相関係数

話題への導入として、鈴木先生の、この好著の、 p. 96 - 101 のおさらいを
する。

生徒 7 名の国語の成績 x と、数学の成績 y の点数が、下表 9.3 の如くある。

表 9.3

生徒番号	P1	p2	p3	p4	p5	p6	p7
国語 x	50	60	65	75	35	80	70
数学 y	40	35	80	55	25	70	45

ピアソン Pearson の 積率相関係数 r を計算して見れば

$$r = 0.7136$$

である。

問題は、国語 と 数学 の教師が別人である為、例えば
生徒 P1 と P2 は、国語と 数学の成績の傾向が異なる。
積率相関係数 r の云うところは、何処まで信頼出来るだろうか？

そこで、成績の順番で、順位 (低から高へ、昇順で) を付けて見る。

表 9.4

生徒番号	P1	p2	p3	p4	p5	p6	p7
x の 順位	2	3	4	6	1	7	5
y の 順位	3	2	7	5	1	6	4

この時、点数値 を 順位値で置き換えた相関係数、即ち

スピアマンの順位相関係数 R は

$$R = 1 - 6 / (n(n^2 - 1)) \sum_{i=1}^n d(i)^2$$

ただし、 $d(i)$ は、生徒 $P(i)$ の x, y の順位の差 difference(i) である。それは、国語の順位で並べ直した表 9.5 で示される。

表 9.5

生徒番号	P5	p1	p2	p3	p7	p4	p6	
国語順位 (昇順)	1	2	3	4	5	6	7	
数学順位		1	3	2	7	4	5	6
差 d	0	-1	1	-3	1	1	1	
$d(i)^2$	0	1	1	9	1	1	1	

$$\begin{aligned} \text{平方和} &= 14 \\ n &= 7 \end{aligned}$$

$$\therefore \text{スピアマンの順位相関係数 } r_s = R = 0.75$$

● 最大の問題： この値は、意味として、信頼出来るか？

鈴木先生の説明： 順位を示す 1 から 7 までの数値には、偏りが無いであろうか？ 元来、ランダムネスの性質を持つ筈である。何故か？ このデータ、7 数字の順位の可能性は $7!$ (7 の階乗) = 5040 通りある。(従って、各順列の生起する確率は $1/5040$ と云う僅かなものである。)

さて、上で計算した「差 d の平方和」が小さい程、相関係数は 1 に近くなる。上の例では 14 であった。

「差 d の平方和」 ≤ 14 と云うような、順位の順列が生起する総数を算えてみれば、全部で 167 通りである(鈴木先生の説明無し)。それ故、今のデータは

$$167/5040 = 0.033 \text{ の確率でしか生起しない筈である。}$$

それが、起きているのだから、これは、信頼すべきである。

と云う訳であるが、問題は 167 通りなる数値を、どうやって

求めたか？ である。この大問題、後節で、私の計算を紹介する。(鈴木先生は、この計算の厄介なことが、

スピアマン法の欠点であると、補足して居られる。)

3. ケンドール の 順位相関係数

順位の逆転の有無に着目した。

国語の順位系列 JJ = 1 2 3 4 5 6 7

数学の順位系列 MJ = 1 3 2 7 4 5 6

両系列の差 JJ - MJ = 0 -1 1 -3 1 1 1 > 0

正なる点検結果は 0 0 1 0 1 1 1

即ち、正 が 4 件 (負は 2 件) ある。

これを、正負逆転の 対 (Tie) T の数と云う。

今は、T = 4 である。

それを使って、

ケンドールの順位相関係数 $r_k = \text{Tau} \text{ タウ} =$
 $1 - 4T/n(n-1) = 1 - 4*4/7*6 = 0.62$ となる。

かくて、ピアソン、スピアマン、ケンドール 等、
3種の相関係数がえられ、互いに、近い値である。

4. 一 致 係 数

今までの議論は、国語と数学と云う如く、2組の変量
について、議論したが、これを拡張して、3組以上、一般に
k組の変量の順位について、成立する尺度が欲しくなる。

これを、ケンドールの「一致係数」と云う。

面倒くさい話を、鈴木先生簡単に処理して居られる。

表 9.4 から 国語の順位 JJ = 2 3 4 6 1 7 5、
,,,,,, 数学の順位 MJ = 3 2 7 5 1 6 4。

複数の順位の和 SJM = JJ + MJ = 5 5 11 11 2 13 9 。

その平均 M = (+/SJM) % n -> 今は 8。

その標準偏差 V = (+/(SJM -M)^2) % n -> 今は 14。

変数の種類 $k =$ 今は 2。

$Mm = k * (n+1) \% 2 \rightarrow$ 今は 8。

$Vm = (k^2) * (n-1) * (n+1) \% 16$ 。

故に、一致係数 $W = V \% Vm = 14 \% 16 \rightarrow 0.875$

良い値である。

鈴木流に、簡単にやれば、こんなところであろうか。

5. 「167 通り」の計算

鈴木先生の説明が無いので、私が代って計算したようなものだ。

いや、スピアマンに代わって、我らが J 言語が計算したと云うのが正解かな？

鈴木先生・北野先生の好著「J 言語による数学計算」の「順列・組み合わせ」の章を活用した。ただし、順列の数値の原点が 0 から始まることは止むを得ない（一々断らぬ）。

```
perm=:?~
plist=: (i. &!A. i.)
posn=: A.

p7 =: plist 7 (=: に注意)
$ p7 -> 5040 (行数、データ総数)
```

先頭行 は $i.7 = 0 1 2 3 4 5 6$

```
p0
0 1 2 3 4 5 6
```

```
7{.p7 (配列 p7 の先頭の 7 行分の例)
0 1 2 3 4 5 6
0 1 2 3 4 6 5
0 1 2 3 5 4 6
0 1 2 3 5 6 4
0 1 2 3 6 4 5
0 1 2 3 6 5 4
0 1 2 4 3 5 6
```

鈴木データ 順列 1 3 2 7 4 5 6 は 配列 p7 中の何番目か？

```
posn 0 2 1 6 3 4 5
138 (それの、この行目に対応)
```

p0 と 後続行との差を採り、その各項の平方和を計算し、その値が ≤ 14 を充たす場合は、該当するとして勘定に入れる。その該当する行の個数を集計する。

スタート行番号は 0 で、最後の行番号は 5040 相当 である。

```
5040 d2cal 0
```

```
start 0
```

```
end 5040
```

```
i,Σd^2 =  
0 0
```

```
i,Σd^2 =  
1 2
```

```
i,Σd^2 =  
2 2
```

```
i,Σd^2 =  
3 6
```

```
i,Σd^2 =  
4 6
```

```
i,Σd^2 =  
5 8
```

```
i,Σd^2 =  
6 2
```

```
i,Σd^2 =  
7 4
```

```
i,Σd^2 =  
8 6
```

```
i,Σd^2 =  
9 12
```

```
i,Σd^2 =  
10 10
```

```
i,Σd^2 =  
11 14
```

```
i,Σd^2 =  
12 6
```

```
i,Σd^2 =  
13 10
```

```
.....
```

```
.....
```

(後段)

```
i,Σd^2 =  
1466 12
```

```
i,Σd^2 =  
1488 14
```

```
i,Σd^2 =  
1560 8
```

```
i,Σd^2 =  
1561 10
```

```
i,Σd^2 =  
1562 10
```

```
i,Σd^2 =  
1563 14
```

```
i,Σd^2 =  
1564 14
```

```
i,Σd^2 =  
1566 10
```

```
i,Σd^2 =  
1567 12
```

```
i,Σd^2 =  
1568 14
```

```

i,Σd^2 =
1572 14
i,Σd^2 =
1584 14
i,Σd^2 =
2160 12
i,Σd^2 =
2161 14
i,Σd^2 =
2162 14
i,Σd^2 =
2184 14
i,Σd^2 =
2280 14 (この後には、該当データは存在せず。)
j =
167

```

集計結果、該当個数は 167 であった。 目出度し！

演算関数 d2cal の script は 後述。

```

d2cal=: 3 : 0 NB. by Y. Nakano 2010/Mar/22
:
wr 'start ', ": y
wr ' end ', ": x

i=.y
j=.0
while. i < x do.
d2 =. +/ ((i.7) - i {P7)^2
if. d2 < 15 do. wr ' i, Σd^2 = '

wr i, d2
j =. j+1 end.

i=.i+1
end.
wr ' j = '
j
)

```

