㊶ RANK CORRELATION METHOD の解説

樋口　伊佐夫

RanK Correlation Method は取扱いが割合簡單で，しかも応用される範囲がかなり広いと思はれる。

それにも拘らず，増山氏の本に少し有る程度で，数理統計学の入門書にはあまり見あたらないので，その大綱を主として Kendall に従つて解説し，初めてこの名に接する方々に，幾分でも参考になることを祈願する。

理論的な事柄や，派生的な問題はすべて省略する。

§ I.　此の方法を適用すべき対象

例えば，色彩の明るさ，硬さ等の様にそれを測る尺度はみつからないが二つのものを比較して何れが他より明るい（硬い）等ということを決定出來る性質がある。（比色分析硬度等）

数個のものゝそういう種類の性質の関係はたゞ順位で表現するより仕方がない。

こういう順位で表現しなければならない二種の性質の間の相関又はこういう性質と他の定量的に測定可能な性質との間の相関を考える時には此の方法が採用される。

又実際行おうと思えばある尺度で測定が出來るものでもたゞ順位だ

けを問題にする場合，例えば体操の時生徒を脊の高さの順にならべようという場合は一一正確な測定をしない。脊が高いということゝ走るのが速いということの間の相関を考える場合，速く走るということは大雑把な概念で正確に規定することは困難である，しかし生徒をみていると客観的な順位というものはある。

一般に脊が高いものは走りにたけているかという大雑把な知識を得るには順位と順位との相関を考えれば十分である。

又つけられる順位は十分客観的なものでなくてもよい。例えば10人の生徒を知性について二人の教師に順位をつけさせて；この二人の教師の判断の間の相関の度合を知ろうという場合にも適用出来る。

しかしこの場合一般に知性というものは linear な関係で測られる性質であるかどうか疑問である。 或る程度この疑問に答える一つの方法を後（§ XIV）に紹介する。

又主として経済学方面であらわれることがあると思われるが，例えば数個の会社の資本金と従業員数との相関を出そうという場合，そのうちの一二が他に比べて非常に大会社である場合は数値をそのまゝつかつてはこの大会社の影響が全体を支配してしまう。

そういう時には資本金の大さの順位と従業員数の順位との相関を考えると或る意味で全体に Balance を與える.

この場合何が問題の要点かということが決定的な事柄であるが常識的に考えて一二の会社の影響しかあらわれない様なのは間違つたやり方である.

§ II. 順位をつけること。

考察の対象となる n 個のものゝあつまりを M とする， Mの要素に普通 1，2，3，-------，n なる自然数を対応させて順位をつける。 しかしこれは本質的な事柄ではない。順位は順序数であつて計量数ではない。

計量数の如く，加えたり，引いたり，乗じたりすることもあるが，その操作が何を意味するか正確に理解することはかなり重要である

M の要素に或る性質に従つて順位をつけようという際，そのうち数個のものゝ間では順位をつけることが出来ない場合が屢々起る。（例えば競走に際し２，３等及び５，６，７等が区別がつかないといつた場合）。

これに順位をつける場合は一応ずらりと通しにつけておいて更めて兄たり難く弟たり難きものについてはその順位の平均を各々につけるという風にする。１０人の競走の時前述のようなことが起つたとすれば，

$$\frac{1}{2}(2+3)=2\frac{1}{2}\ ,\quad \frac{1}{3}(5+6+7)=6$$

であるから

$$1,\ 2\frac{1}{2},\ 2\frac{1}{2},\ 4,\ 6,\ 6,\ 6,\ 8,\ 9,\ 10$$

と順位をつける。

こういう順位づけ（ranking）を「括られた順位」（tied rank）をもつ ranking という。又 $2\frac{1}{2}$ 及 6 の様に同じものがある順位を「括り」（tie）といいその数を tie の拡がりという。即ち上の ranking は拡がりが２と３の二つの tie をもつている。

§ III　二つの Ranking の間の相関を測る測度

先づ tie のない場合を考える。

M の要素を A，B，C 等であらわし，第一の順位づけによる順位を夫々 p_A，p_B，p_C 等であらわし，第二の順位づけによる順位を夫々 q_A，q_B，q_C 等であらわす。

二つの順位づけの間の相関の測度を次の基準で選ぼう。

1）二つの順位づけが完全に一致している時即ち M の任意の要素 A に対して $p_A=q_A$ なる時 ＋1を與える。

2）一方の順位づけを逆にすると他方と全く一致する場合，即ちM の任意の二要素 A，B に対して $p_A>p_B$ ならば $q_A<q_B$ なる時，

−1 を與える。

3) 其の他の場合は +1 と −1 の中間の値をあたえる。但しその際何等かの意味で順位づけの一致の度合に応じて増加してゆく風になつている。

1), 2) の要請は全く習慣的なものではあるが甚だ有用である。

1), 2), 3) の要請をみたすものとして最も自然な考え方は Kendall のいわゆる順位相關 τ である。

それは, $p_A > p_B$ 且 $q_A > q_B$ なる A, B の対の数を P とし $p_A > p_B$ 且 $q_A < q_B$ なる A, B の対の数を Q とし, M のあらゆる対の数を T とする時

$$\boxed{\tau = \frac{P-Q}{T}} \quad \cdots\cdots\cdots\cdots (1)$$

で定義する。 この τ は tie のない場合はあきらかに上記 1) 2), 3) の性質をもつている。

一つの例を考えよう。

	A	B	C	D	E	F	G	H	I	J
I	9	2	8	6	4	3	5	1	10	7
II	3	6	9	4	8	2	5	7	10	1

例えば (C, E) の対については I での順位数も II での順位数も共に C の方が大きい。 こういう対に対しては +1 を記録する。

又 (B, C) の対に対してもともに C の方が大きいから +1 を記録する。 然し (H, F) の様に I での順位数は F の方が大きいが II での順位数は H の方が大きいものには −1 を記録する。

この 10 個のものについて対は ${}_{10}C_2$ 個出來るがこの 45 個の対について記録を合計したもの S が P−Q である。

この例では S = +1 となり

$$\tau = \frac{+1}{45} = +0.02 \quad である。$$

tie のない場合では $P+Q=T={}_nC_2$ だから

$$\tau=\frac{P-Q}{{}_nC_2}=\frac{2P}{{}_nC_2}-1=1-\frac{2Q}{{}_nC_2}\cdots\cdots\quad(2)$$

次に Spearman の ρ を説明する。

上の例を再び書く。

	A	B	C	D	E	F	G	H	I	J
I	9	2	8	6	4	3	5	1	10	7
II	3	6	9	4	8	2	5	7	10	1
差 d	6	-4	-1	2	-4	1	0	-6	0	6
d^2	36	16	1	4	16	1	0	36	0	36

d^2 の和を $S(d^2)$とし

$$\boxed{\rho=1-\frac{6\,S(d^2)}{n^3-n}}\cdots\cdots\cdots\cdots\cdots\cdots(3)$$

と定義する。この ρ も前記 1), 2), 3) の性質をもつ。

この例では $\rho=+0.12$ である。

上の例で次の様に書きなおしてみる。

	H	B	F	E	G	D	J	C	A	I
I	1	2	3	4	5	6	7	8	9	10
II	7	6	2	8	5	4	1	9	3	10

こうしても明かに τ や ρ の値に変化はない。だから一方が自然の順序になつている様にならべかえることが屡々ある。

言うまでもないことであるが τ を勘定する時，例えば（F, I）（F, G）の二つの対についてみるに同じく＋1を記録するのであるが，Iについての順位の差は夫々7，2であり II についての順位の差は

8．3である。

それに同じ1を対応させるのは変な気がするが 順位というものは或る尺度で測定された測定値ではない。 1，2，----nでつけた時二個の要素の順位数の差とはその二つの要素の間に入つてくるところの要素の数（に1を加えたもの）であることを理解すべきである。

例えばある尺度で測定出来たとして順位が5のものと10のものの間は順位が3のものと5のものとの間に比して100分の1であるかも知れない。

Rank の理論の歴史的な起りは測定する尺度をつくりあげたり複雑な計算をしたりする面倒さをはぶくために実用的に近似方法として考えられた。 それ故 rank を variate と見て来た。その限りでは Spearman の ρ の方がすぐれている！

しかし順位を保存する様に P_A 等の値をかえてみる。すると τ は何等影響をうけないれど ρ は忽ち値を変ずる。ordinal number を基にして ρ の意味を考えると甚だ面倒になる！

§IV. τ の 勘 定 の 仕 方

Tieのない場合は，（2）の式により結局 P 又は Q を勘定すればよい。 簡単に P を勘定する方法を述べる。 前節に述べた如く一方が自然の順序になる様にならべ換えておく。 他方の順位づけに於ける順位数に着目して，そのものより右にあり，且つそのものより順位数の大きいものの個数を順々に加えて行く。 その和が P，である。 前例で説明すると第Ⅱの順位づけに着目する。 例えば7の右にあつて7より大きいものは8，9，10 の3個である。次に6の右にあつて6より大きいものもやはり8，9，10 の3個である。

次に2 の右にあつて 2 より大なるものは 8，5，4，9，3，10，の6個である。 同様にして結局 Q＝3＋3＋6＋2＋2＋2＋3＋1＋1＝ 23 を得る。

もしも何れか一方が自然の順序になる様ならべかえるのが面倒ならば，自然の順序を上に書いておく。　そして次の様にする。

	1	2	~~3~~	4	5	6	~~7~~	8	9	10
I	9	2	8	6	4	3	5	1	10	7
II	3	6	9	4	8	2	5	7	10	1

先づ II の順位づけに於て１のものから始める。　そのものの I での順位は７である自然の順序に於て７の右側には３個ある。

先づ３と記録し自然の順序に於ける７を消しておく。　次に II の２に対応する I は３である。　自然の順序に於て３の右側にあるものは６個である。（７は既に消されているから）。　それで６と記録して自然の順序に於ける３の数字を消しておく。　同様の事を；II で３であるもの，４であるもの等につき順々に行う。　そして記録を合計してＱを得る。

$$Q = 3+6+1+2+2+3+3+2+1 = 23$$

§ V. 一般相関係数とτ及びρ

便宜上 M の要素に 1,……, n と番号をつけて置く。（これは順位ではなく単なる呼び名である）これから M の要素をこの番号で呼ぶことにする。

M の各要素が持っている二種の性質 x, y に関して考察する。

M の任意の二つの要素例えば i, j に対し性質 x に関する何らかの数値 a_{ij} があるものとする。それにはたゞ $a_{ij}=-a_{ji}$ なる條件が附されているとする。

又性質 y に関して b_{ij} があつて $b_{ij}=-b_{ji}$ をみたすものとする。

従つて $a_{ij}=b_{ij}=0$ ($i=j$)

一般相関係数 Γ は次の式で定義される。

$$\boxed{\tau = \frac{\sum a_{ij} b_{ij}}{\sqrt{\sum a_{ij}^2 \sum b_{ij}^2}} \quad (\textstyle\sum \text{ は } \sum_{i=1}^{n} \sum_{j=1}^{n} \text{ をあらわす})} \qquad \cdots\cdots (4)$$

τ は，

$$\left.\begin{aligned} a_{ij} &= 1 \quad (p_i > p_j \text{ なる } i,j \text{ に対して}) \\ &= -1 \quad (p_i < p_j \text{ なる } i,j \text{ に対して}) \\ b_{ij} &= 1 \quad (q_i > q_j \text{ なる } i,j \text{ に対して}) \\ &= -1 \quad (q_i < q_j \text{ なる } i,j \text{ に対して}) \end{aligned}\right\} \cdots\cdots (5)$$

としたものであり，

ρ は

$$\left.\begin{aligned} a_{ij} &= p_i - p_j \\ b_{ij} &= q_i - q_j \end{aligned}\right\} \cdots\cdots\cdots\cdots (6)$$

としたものであり，

更に普通用いられる積率相関係数即ち

$$\frac{cov(x, y)}{\sqrt{var\, x\; var\, y}} \quad \text{は}$$

$a_{ij} = x_i - x_j \quad b_{ij} = y_i - y_j$ と実際の variate の値の差を a_{ij}，b_{ij} とした時の一般相関係数であることは簡単な計算の結果わかる。

§ VI　τ 及び ρ の拡張

先づ（1）を tie のある場合にもそのまゝ用いる拡張の仕方がある。その際言うまでもなく $p_A = p_B$ 又は $q_A = q_B$ なる対 A,B に対しては 0 と記録する。だからもはや（2）の式は成立たない。

これを τ_a であらわす

$$\boxed{\tau_a = \frac{S}{T}} \quad \cdots\cdots\cdots\cdots (1')$$

次に（4），（5）を見ると新しく $a_{ij}=0$（$p_i=p_j$），$b_{ij}=0$（$q_i=q_j$）という至極妥当なる條件を附加えることにより tieのある場合にも拡張出來る。 その場合 $\Sigma a_{ij}b_{ij}$ は（1）の場合のSの２倍になつている。 たゞ分母が少し異る。

Ⅰ にふくまれる tie の拡がりを u とし $\frac{1}{2}u(u-1)$ をつくるそれを Ⅰ に含まれる すべてのtieについて加えたものをUとする。又Ⅱに含まれるtie…… tieのひろがりを v とし Ⅰ の場合と同じく V をつくる 即ち

$$\left.\begin{aligned} U &= \frac{1}{2}\sum_u u(u-1) \\ V &= \frac{1}{2}\sum_v v(v-1) \end{aligned}\right\} \quad \cdots\cdots\cdots\cdots (7)$$

とすれば，此の拡張は

$$\frac{S}{\sqrt{(T-U)(T-V)}}$$

となるこれを τ_b とする。即ち

$$\boxed{\tau_b = \frac{S}{\sqrt{({}_nC_2-U)({}_nC_2-V)}}} \quad \cdots\cdots\cdots\cdots (8)$$

常に $\tau_b \geqq \tau_a$ である。 τ_a は一般に § Ⅲ であたえた基準 1），2）をみたさないけれど，τ_b は 1），2）をみたす。

τ はもつと拡い意味での順位づけに対して拡張出來るがこゝでは立入らない。

τ_a と τ_b は目的によつて夫々使いわける。 例えば二つの判断の一致の度合を測るには，τ_b を用いた方がよい。 何故なら二つの判

断（順位づけ）が完全に一致していても tie が相当大きければ τ_a は非常に小さい値を与えるからである。 しかし次の様な目的には，τ_a を用いた方がよい。 即ち例えばちゃんと比色分析で順位のついている数種の青色を洋裁師に見せてその判断による明暗の順位をつけさせて見わける能力をはかろうという場合は τ_b を用いるとよくないことは明かだろう。

Tie を能力があればその間に順位をつけ得る筈のところ，能力がないためにつけ得ないのだと考え，ひろがり t の tie について，t ！の可能な整数の順位を入れ，それらについて τ の平均をとってみるとそれは τ_a になる。

記号は前と同じとして

$$\left.\begin{aligned} \frac{1}{12}\sum_u (u^3-u) &= U' \\ \frac{1}{12}\sum_v (v^3-v) &= V' \end{aligned}\right\} \quad \cdots\cdots\cdots\cdots (9)$$

とすると，tie のある場合えの ρ の拡張は

$$\boxed{\rho_a = 1 - \frac{6\{S(d^2)+U'+V'\}}{n^3-n}} \quad \cdots\cdots\cdots\cdots (10)$$

$$\boxed{\rho_b = \frac{\frac{1}{6}(n^3-n)-S(d^2)-U'-V'}{\sqrt{\{\frac{1}{6}(n^3-n)-2U'\}\{\frac{1}{6}(n^3-n)-2V'\}}}} \quad \cdots\cdots (11)$$

であたえられる。 ρ_a はもしも tie を自然数の tie のない順位におきかえた時あらゆる可能な置き換えに対する ρ の値の平均値であり ρ_b は（4），（6）を拡張したものである。

§ VII. 風変りな應用

心理学ではかなり普通の問題であるが，一見風変りな応用に次の様なものがある。

M の要素をある性質をもつかもたないかによりニつの組に分類する。その分類のことを Dichotomy という。

M に與えられた，順位づけと Dichotomy との間の関係をはかるのに τ_b が応用される。例えば，男の子と女の子が試験の成績によつてならべられた時，性と成績との間に何らかの関係があるだろうかという問題がこれである。

今，成績のいいものから順に 1, 2, 3, と番号をうつ。そして次の様な結果が得られたとする。

1	2	3	4	5	6	7	8	9	10	11	12	13
男	男	女	男	女	女	男	男	女	男	女	男	男

強いて順位をつけるならば，男の順位を女の順位より小さくつけた場合 τ_b が $+1$ に近ければ男は成績がいいということなり，-1 に近ければ女の方が成績がいいということになる。

τ を考える限りは男の順位数を $(1+2+3+4+5+6+7+8)\div 8 = 4\frac{1}{2}$，女の順位数を $(9+10+11+12+13)\div 5 = 11$ にしなければならぬ理由はさらにない。

§ VIII. 無相関の検定

我々は M の要素を二つの性質 x, y に従つて順位をつけて，その順位の間の相関によつて M の x, y の間の関係を知る一つの手段にすることが出来る。次に M が，M を含む更に大きな母集団からの，一つの標本であると考え，母集団に於ける性質 x, y の関係を知りたい。先ず母集団に於ては x と y とは関係がないと仮

定する。 そういう仮定の下では標本は tie があつたり，なかつたり，種々の場合があるが，今標本値が tie をもたない場合ばかりを取立てゝ考えるならば，性質 x に関する或る與えられた一つの順位づけ Ⅰ に対して，性質 y に関するあらゆる可能な $n!$ 通りの順位づけは，任意標本である限り，すべて同じ確率を以てあらわれるだろう。

この $n!$ 通りの Ⅱ の順位づけを Ⅰ の順位づけとの間の τ（又は ρ）の値により分類すれば τ（又は ρ）の頻度分布が得られる。 こうして得られた頻度分布はたゞ n のみに依存することを注意しておこう。

次に tie を含む場合は，一定の構造の順位づけばかりとりたてゝ考える。 例えば 1，$2\frac{1}{2}$，$2\frac{1}{2}$，5，5，5，7，$8\frac{1}{2}$，$8\frac{1}{2}$ という順位づけを考える時は，M の要素に，この番号を一応すべて異るものと考えて，対応させるさせ方は 9! 通りあるが，その各々が同じ確率であらわれると考える。

又これとは独立に 2，2，2，$4\frac{1}{2}$，$4\frac{1}{2}$，6，7，8，9 なる第Ⅱの順位を対応させる 9! 通りの仕方は第Ⅰの順位づけに無関係にすべて同じ確率であらわれると考える。

第Ⅰの順位づけを一つ固定しておいて第Ⅱの順位づけを $n!$ 通り動かして得る τ（又は ρ）の頻度分布は第Ⅱの順位づけを一つ固定しておいて第Ⅰの順位づけを動かして得る τ（又は ρ）の頻度分布と同じものであることを附言しておこう。

これ等を用いて相関がないという仮説に対する検定を行うことが出来る。 τ の分布の代りに S の分布を考える。それに関する知識は次の通りである。 tie のない場合.

分布は常に $S=0$ に対して対稱である。即ち平均は0である。

又もしも $\frac{1}{2}n(n-1)$ が偶数ならば S は常に偶数だけをとり，$S=0$ に於ける頻度は最大である。

$\frac{1}{2}n(n-1)$ が奇数なら S は奇数だけをとり $S=\pm1$ で最大である。何れにしても頻度は Maximum から $S=\pm\frac{1}{2}n(n-1)$

（こゝでは頻度は1）になる所を順々におちてくる。　つまりあらわれる S の値に対する頻度の点をつないで得られる Graph は山が一つである。そして

> n が増加するにつれて S の頻度分布は平均 0，標準偏差 σ の正規分に近づく。　σ は S の分布の標準偏差で
>
> $$\sigma^2 = n(n-1)(2n+15)/18 \cdots\cdots (12)$$
>
> で與えられる。

$n > 10$ なる時はこの近似は十分有効である。　$n = 10$ までは表を附しておいた。　$n > 10$ に対しては正規分布の表を用いていただきたい。　但しその際次のことに留意されたい。　とり得る値の間隔が2である所の離散分布を連続分布で近似したのだから尾の部分を考慮すると測定された S が正の場合は S のかわりに S−1 を用い，S が負の場合は S の代りに S+1 を用いた方が近似がよくなる。

tie のある場合の正確な分布は種々な場合によつて個々に計算されねばならない。　又正規分布で近似する時 variance を次の様に修正しなければならない。

$$\sigma^2 = \frac{1}{18}\Big\{n(n-1)(2n+5) - \sum_u u(u-1)(2u+5) - \sum_v v(v-1)(2v+5)\Big\}$$
$$+ \frac{1}{9n(n-1)(n-2)}\Big\{\sum_u u(u-1)(u-2)\Big\}\Big\{\sum_v v(v-1)(v-2)\Big\}$$
$$+ \frac{1}{2n(n-1)}\Big\{\sum_u u(u-1)\Big\}\Big\{\sum_v v(v-1)\Big\} \cdots\cdots (13)$$

一方にしか tie がないとか dichotomy の場合はこの式は容易にもつと簡單にすることが出来る。

例えば，一方が dichotomy の時は $v = s$ と $v = r$ の二つの tie があると考えて計算すると

$$\sigma^2 = \frac{r\,S}{3n(n-1)}\left\{n^3 - n - \sum_u (u^3 - u)\right\} \cdots\cdots (14)$$

となる。 其の他の場合も容易だから省略する。たゞ連続分布で近似した場合の S に対する補正に関して次の注意をしておく。

a) 何れか一方の順位づけが tie を含まず他方は dichotomy である場合 S の可能な値の間隔は2であるから補正は +1.

b) 何れか一方が ~~一方が~~ 拡がり t の tie ばかりからなり他方が dichotomy である場合は間隔は 2t であるから補正は t.

c) 何れも dichotomy の時は間隔は N だから補正は $\frac{1}{2}$N.

d) 一方が dichotomy で他方がいろいろの大きさの tie を含む時 S の間隔は S の値のいろいろの部分により異るが近似的方法がある。次の例でそれを説明しよう。

例 或る講習会が終つた後，受講者の幾人かに講習会が有意義であつたか否かをたづねた。出席回数の多いものは有意義であつたと答えるかどうか。出席率の最高のものを1として出席率に従つて受講者に順位をつければ次の様な結果を得た。

有意義と答えた者 $1\frac{1}{2}$, $1\frac{1}{2}$, 3, $6\frac{1}{2}$, $6\frac{1}{2}$, 9, 12, 12, 12, 17.

無意義と答えた者 4, $6\frac{1}{2}$, 12, 12, 15, 16.

$S = 23$, $\tau_b = 0.26$ を得る。(14)により σ^2 を計算すると

$$\sigma^2 = \frac{6\times 11}{3\times 17\times 16}\left\{(17^3-17)-(2^3-2)-(4^3-4)-(5^3-5)\right\}$$

$$= 341.47$$

$$\sigma = 18.48$$

扨，話をすゝめるために順位づけのある方（dichotomy でない方）を大きさの順にならべる。

$1\frac{1}{2}, 1\frac{1}{2}, 3, 4, 6\frac{1}{2}, 6\frac{1}{2}, 6\frac{1}{2}, 6\frac{1}{2}, 9, 12, 12, 12, 12, 12, 15, 16, 17.$

S の間隔の和は

2n－（順位数の最小のものを含むtieのひろがり）

－（順位数の最高のものを含む tieのひろがり）

で與えられる。

我々の場合は17は tie ではないがひろがり1の tieと考えて

$$2 \times 17 - 2 - 1 = 31$$

そしてこれを順位のとびの場所で平均する。

我々の場合は $1\frac{1}{2}$ と 3，3と4，4と$6\frac{1}{2}$，$6\frac{1}{2}$と9，9と12，12と15，15と16，16と17 の 8 箇所あるから $31/8 = 3.875$ これが S の間隔の平均であることが証明出來る。 それ故補正としてはその半分 1.937 を用いる。

$$\frac{23 - 1.937}{18.48} = 1.14$$

1.14は N(0,1) （平均0，標準偏差1の正規分布）の36%点に相当するからこれは有意ではない。 故にこの data からでは関係があると結論を下すわけにはいかない，

§ IX. ρ による検定

前節に説明した ρ の分布は原点に対して対稱ではあるが τ の時の様にはすみやかに正規には近づかない。 然し tie があつてもなくても常に

$$var\,\rho = \frac{1}{n-1} \quad \cdots\cdots\cdots\cdots \quad (13)$$

である。 $n>20$ ならば正規分布を用いることが出来る。

$n \geqq 9$ では

$$t = \rho\sqrt{\frac{n-2}{1-\rho^2}} \quad \cdots\cdots\cdots \quad (14)$$

が自由度 $n-2$ の Student t—分布に近似的に従うとしてよい。

$n \leqq 8$ に対しては ρ の分布の表は出来ているが計算方法を示すにとどめる。

$$\begin{vmatrix} a^0 & a^1 & a^4 & \cdots\cdots & a^{(n-3)^2} & a^{(n-2)^2} & a^{(n-1)^2} \\ a^1 & a^0 & a^1 & \cdots\cdots & a^{(n-4)^2} & a^{(n-3)^2} & a^{(n-2)^2} \\ a^4 & a^1 & a^0 & \cdots\cdots & a^{(n-5)^2} & a^{(n-4)^2} & a^{(n-3)^2} \\ \cdots & \cdots & \cdots & \cdots\cdots & \cdots & \cdots & \cdots \\ a^{(n-2)^2} & a^{(n-3)^2} & a^{(n-4)^2} & \cdots\cdots & a^1 & a^0 & a^1 \\ a^{(n-1)^2} & a^{(n-2)^2} & a^{(n-3)^2} & \cdots\cdots & a^4 & a^1 & a^0 \end{vmatrix}$$

なる行列式を考える。 これを，負号をつけない様に展開する。その意味は普通行列式を $n!$ 個の項に（各項を同じ行及び同じ列に含まれることのない様な n 個の要素の積として）展開する際のすべての負号を正号になおして和をつくる。そうした時の a^m の係数が，$m = S(d^2)$ の値に対する頻度である。

§ X. 母相関のある場合の検定

母集団で相関がないという仮説が捨てられた場合相関はどれくらいかという知識を何らかの確率的な意味ででも得たい。

そういう目的の為に§VII とは異った見地から検定をつくる。

先づ大きさ N の母集団 $\mathfrak{M}$ からの大きさ n の標本が M である

と考える。

$\mathfrak{M}$ の要素はすべて x 及 y の性質に関してきまつた順位をもつているとと考える。 $\mathfrak{M}$ からとりだす標本に従つて必然的に τ の標本値はきまるのである。

母相関を τ として τ の標本値を t であらわす。 ${}_N C_n$ 通りの size n の標本に対する t の平均を $E(t)$ とすると常に

$$E(t) = \tau \quad \cdots\cdots\cdots\cdots\cdots\cdots \quad (15)$$

この ${}_N C_n$ 通りのとり方に対する t の値の分布は n, N がともに増して行つた時，τ が ± 1 に近くない限り平均値 τ の正規分布にだんだん近づいてゆく。 その際 variance が τ のみによることが望ましいのだがそういうわけにはゆかない。 しかし次の関係があるので安全な側の検定をあたえる。

$$\boxed{var\, t \leqq \frac{2}{n}(1-\tau^2)} \quad \cdots\cdots\cdots\cdots \quad (16)$$

例えば，２０のものの順位づけに於て τ の値０.７８９を得た時，この順位づけを任意標本とみるときは，母集団の τ の値に関してどういうことが言えるか？ $n=20$ に対しては t の分布を正規としてよい。 そして t の平均値の推定値として０.７８９をとると，(１６)から標準偏差は０.１９４より大きくはないことが出る。 故に正規分布の性質から ０.７８９±(０.１９４×１.９６) の中には真の τ があるということは ９５％ 以上の確さをもつていえる。

τ は１より大きいことはないから，τ は０.４０９から１までの中にあると言える。 先に順相関がないにも拘らずあると結論する危険はない。そういう意味で安全な側の検定をあたえる。 又信頼区間の理論をつかえば未知の τ を標本値 でおきかえる時に生ずる 誤差の危険をなくすることが出来る。 N (０.１)の P パーセント水準点を x_p とすると τ の限界値は

$$\tau = \frac{t \pm x_p \sqrt{\frac{2}{n}\left(1+\frac{2}{n}x_p^2 - t^2\right)}}{1 + \frac{2x_p^2}{n^2}} \quad \cdots\cdots\cdots\cdots\cdots\cdots (17)$$

となる。　前の例では $\tau = 1.3$ しき と 0.240

即ち　τ は 0.240 と 1 の間にあるということになる。

$var\ t$ を（１６）より小さくおさえることは不可能であることが証明されている。それ故母相関を推定するには n が３０から４０より大きい場合でないと幅がひろがりすぎて目的にかなわない。

だから数個の標本があわせ考えられるのでなければrankから勘定した相関係数の値で母集団に関して云々するのには十分の注意を要する。

（５）に於ける a_{ij} b_{ij} に対して

$$a_{ij}\,b_{ij} = c_{ij}\ \sum_{j=1}^{n} a_{ij}\,b_{ij} = c_i\,,\ \sum_{i=1}^{n} c_i = c$$

とすれば $c,\,c_i,\,c_{ij}$ の標本値に対して $var\ t$ の不偏推定量は

$$var\ t = \frac{1}{n(n-1)(n-2)(n-3)}\left\{4\sum c_i^2 - \frac{2(2n-3)}{n(n-1)}c^2 - 2n(n-1)\right\}$$

$$\cdots\cdots\cdots\cdots\cdots\cdots (18)$$

であたえられることを利用すれば少し幅をせばめることが出来ることもある。

更に標本分布が正規型からはなれていることを考慮して

$$\gamma_1 = E(t^3) \Big/ \{E(t^2)\}^{\frac{3}{2}}$$ とすると

$$\pm x_p + \frac{x_p^2 - 1}{6}\gamma_1 \quad \cdots\cdots\cdots\cdots\cdots\cdots (19)$$

によって $x = \dfrac{t-\tau}{\sqrt{E(t^2)}}$ の Ｐ パーセント信頼限界が得られる。

Ｅ（t^3）の推定値は

$$E(t^3)=\frac{8}{n^6}\left\{\sum_{i>j} C_{ij}(C_i+C_j)^2-\frac{5C\sum C_i}{n}+\frac{3C^2}{n}\right\}$$

で近似的に与えられる。ということが Kendall と Daniels により得られている。

§ XI. m 個の Ranking がある場合の全体的関係

今 M に対して m 通りの順位がついている場合これ等の順位づけの間の全体としての関係を測る測度と考えよう。

${}_mC_2$ 個の τ（又は ρ）が得られるからその平均をとるということが妥当であろう。

しかしこれは m が大きい時は計算は面倒である。

今 M の要素には（§ V の初めに言った様に）番号がついているものとする。 m個の順位づけは

$$\begin{matrix} r_{11}, & r_{12}, & \cdots\cdots, & r_{1n} \\ r_{21}, & r_{22}, & \cdots\cdots & r_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ r_{m1} & r_{m2}, & \cdots\cdots & r_{mn} \end{matrix}$$

であらわす。 即ち r_{ik} は i 番目の順位づけに於ける対象 k のもつ順位である。

$$\boxed{\sum_{k=1}^{n}\left\{\sum_{i=1}^{m} r_{ik}-\frac{1}{n}\sum_{k}\sum_{i} r_{\cdot k}\right\}^2=S} \quad \cdots\cdots \quad (20)$$

とする。

tie のない場合

$$\boxed{W = \frac{12\,S}{m^2(n^3-n)}} \quad \cdots\cdots\cdots\cdots (21)$$

tie のある場合は各順位づけに於て § VI (7) の意味で夫々，$\frac{1}{12}\sum_u (u^3-u)$ をつくり i 番目の順位づけのものを U_i' とするとき，

$$\boxed{W = \frac{S}{\frac{1}{12}m^2(n^3-n) - m\sum_{i=1}^{m} U_i'}} \quad \cdots\cdots\cdots\cdots (22)$$

と定義する　この W のことを Coefficient of Concordance という。$0 < W \leqq 1$ となる。又 $W=1$ となるのはすべての k に対して $r_{1k} = r_{2k} = \cdots\cdots = r_{mk}$ となる時且つその時に限る。m 個の順位づけから得られる ${}_mC_2$ 個の ρ の平均を ρ_{av} とすると tie のない場合は

$$\rho_{av} = \frac{mW-1}{m-1} \quad \cdots\cdots\cdots\cdots (23)$$

次に m 個の順位づけが独立であるという仮説に対する有意性検定を考える。それには $(n!)$ 個の可能な m — 順位づけがすべて同じ確率であらわれると仮定した時の W の値の頻度分布を求める。

tie のない場合 m と n の小さい部分に対しては正確な分布の表が出来ている。$n \geqq 6$ の場合には次の近似が十分適用される。

$$Z = \frac{1}{2}\log_e\frac{(m-1)W}{1-W} \quad \cdots\cdots\cdots\cdots (24)$$

は自由度 $\nu_1 = n-1-\frac{2}{m}$　　$\nu_2 = (m-1)\nu_1$ の Fisher の Z 分布に近似的に従う

$n \geqq 6$ でなくても $n=5$ なら $m \geqq 4$；$n=4$ なら $m \geqq 6$；

$n=3$ なら $m \geqq 6$ でありさえすれば，十分この近似は有効である。

こゝで自由度が整数でない場合も起るが Z 分布や χ^2 の分布函数於て自由度をあらわす Parameter に実数を入れたものはやはり分布函数の性質をもっている。

自由度が整数でないものに対しては表はないが自由度の逆数に関して内挿法を用いる。

即ち，ν_1，ν_2 が分数の時 Z の1パーセント水準点を求めるには自由度が夫々 $[\nu_1]$，$[\nu_2]$；$[\nu_1]$，$[\nu_2]+1$；$[\nu_1]+1$，$[\nu_2]$；$[\nu_1]+1$，$[\nu_2]+1$ なる分布の1パーセント水準点は表で得られるその値を夫々 a，b，c，d とする

$$\left\{\frac{1}{[\nu_1]}-\frac{1}{\nu_1}\right\}\Big/\left\{\frac{1}{[\nu_1]}-\frac{1}{[\nu_1]+1}\right\}=\theta$$

$$\left\{\frac{1}{[\nu_2]}-\frac{1}{\nu_2}\right\}\Big/\left\{\frac{1}{[\nu_2]}-\frac{1}{[\nu_2]+1}\right\}=\varphi$$

とすると

$$a-(a-b)\varphi-(a-c)\theta+(a+d-c-b)\theta\varphi$$

で與えられる。

最後の項は普通省略して差支えない。

（$[\nu_1]$ 等は ν 等の整数部分をあらわす）

しかし $n>\eta$ に対しては次の近似を用いて大過ない。

$$\boxed{\chi_r^2=m(n-1)W=\frac{12S}{mn(n+1)} \qquad \cdots\cdots\cdots\cdots (25)}$$

が自由度 $\nu=n-1$ の χ^2 分布に近似的に従う

tie のある場合は tie の数やそのひろがりが大きくない時は，Z—検定はそのまゝつかえるがそうでない場合は（２０）の時の U_i' を用い）

$$\left.\begin{aligned} \mu_{2i} &= \frac{1}{12}(n^2-1) - \frac{1}{n}\,\mathrm{U}_i' \\ \mu_2(W) &= \frac{4}{m^2(n-1)}\;\frac{\sum_{i\neq j}\mu_{2i}\mu_{2j}}{(\sum_i \mu_{2i})} \end{aligned}\right\} \cdots\cdots\cdots (26)$$

とおくと（２４）の事柄は

$$\left.\begin{aligned} \nu_1 &= \frac{2(m-4)}{m^3\mu_2(W)} - \frac{2}{m} \\ \nu_2 &= (m-1)\nu_1 \end{aligned}\right\} \cdots\cdots\cdots (27)$$

とすれば用いられる。

又（２５）のかわりに近似的 χ_r^2 には

$$\chi_r^2 = \frac{S}{\frac{1}{12}mn(n+1) - \frac{1}{n-1}\sum_{i=1}^{m}\mathrm{U}_i'} \cdots\cdots\cdots (28)$$

§ XII　順位をestimateすること。

Wが有意であるとするとm－順位づけの間に何らかの一致があることになるがその時一つの眞の順位づけがあつてm通りの値が得られたのは偶然的契機によるものであると考え眞の順位づけをestimateすることを問題にすることが出来る。

その場合簡単なのは ρ による最小自乗法に基礎をおくものでm－順位づけの和に従つて順位をつける。　つまり $\sum_{i=1}^{k} r_{ik}$ による。

X	Y	Z
5	6	9
4	5	7
10	5	2
2	5	3
21	21	21

然し例えば，４通りの順位づけに際し３個の対象X，Y，Zに関し左の様になつたとする。

眞の順位づけにtieをゆるすならばこれは同じ順位をつけてよいが眞の順位づけにはtieがないという前提の下では平方和をと

ると，X，Y，Z，は夫々 145，111，143 になるからY，Z，Xの順に順位をつける。 tie が m 順位づけの中にあつてもこの方法は用いられる。

相関のある場合の W の検定に対する分布は殆んど知られていないために，いくつかの順位づけの組について和合度の差があるという仮説は検定出来ない。 例えば20人の生徒のある性質について男の教師7人と，女の教師5人が順位をつけた際男の教師の間では W＝0.5 女の教師の間では W＝0.7 であつたとする。このことから一般に女教師は男教師に比してこのことに対してより一致した批判を下すものであると結論してよいかどうかということは残念ながら今の所知るすべがない。

§ XIII 偏順位相関

普通の偏相関係数の様に，P，Q，R，の三つの順位づけがある時 Q と Rとの間の τ を τ_{QR} という風に書いて

$$\boxed{\tau_{QR\cdot P} = \frac{\tau_{QR} - \tau_{PR}\tau_{QR}}{\sqrt{(1-\tau_{PR}^2)(1-\tau_{QP}^2)}}} \quad \text{·············(2.9)}$$

で Q と R の間の偏順位相関を定義する。

		順位づけ R		和
		Pと同じ向きの対の数	Pと逆向きの対の数	
順位づけQ	Pと一致する対の数	a	b	$a+b$
	Pと逆向きの対の数	c	d	$c+d$
和		$a+c$	$b+d$	$N=(a+b+c+d)={}_nC_2$

とわけると

$$\tau_{QR} = \{(a+d)-(b+c)\}/N$$

$$\tau_{PQ} = \{(a+b)-(c+d)\}/N$$

$$\tau_{PR} = \{(a+c)-(b+d)\}/N$$

であるから

$$\tau_{QR.P} = \frac{ad-bc}{\sqrt{(a+b)(c+d)(a+c)(b+d)}} \quad \cdots\cdots\cdots\cdots (30)$$

となる。 これは明らかに $|\tau_{QR\cdot P}| \leqq 1$

$|\tau_{QR\cdot P}| = 1$ ならば a, b, c, d ,の中少くとも二つが0であるということになる。

もしも $a=b=0$ 又は $c=d=0$ ならば $|\tau_{PQ}|=1$

$a=c=0$ 又は $b=d=0$ ならば $|\tau_{PR}|=1$ となり意味はなくなる。

それで $a=d=0$ 及び $b=c=0$ の時を考える。

前者の場合は $\tau_{QR.P}=-1$ 後者では $\tau_{QR.P}=+1$

場合は必然的に夫々 $\tau_{QR}=-1$ $\tau_{QR}=1$ となるこの……

又この $\tau_{QR.P}$ は Q と R との相関は真に P と Q 及び P と R の相関に依存しているかどうかという知識を或る程度與える。

即ち $\tau_{QR\cdot P}$ は P の影響に無関係な Q と R の間の一致性に対する一つの尺度を與えているということが出来る。

例えば, $a/b = c/d$ ならば Q と R は, それ等が共に P と一致していなければ, 一致しているとはいい難い。

R に於て P に於けると同じ向きの対の数と逆向きの対の数との比が Q で P と一致しているものと一致しないものと等しいからである。 それ故, P と一致しようとしまいと Q と R が一致すればするほど偏相関は増大する。

例えば,

Ⅰ	1	2	3	4	5	6	7	8	9	10
Ⅱ	2	3	1	5	6	7	4	10	8	9
Ⅲ	4	1	2	3	6	5	8	7	9	10

を見ると，Ⅰ，Ⅱ，の関係と，Ⅰ，Ⅲ，の関係が，Ⅱ，Ⅲ の間の関係におおいをかけていると感づく。

しかしこの種の事柄は十分遠慮して言わねばならない。

偏-ていに関する検定も未だ知られていない。

§ XIV. 順位をつけることが妥当かどうか。

最初に提起して残して来た問題を考える。

即ち，Mにある性質に関して順位をつける際，その性質は果して一般に順位づけが出来るのかどうかという問題である。

これは M の特性でなくて性質そのものに本質をもつ。この問題に答えるためにたまたま M が選ばれる。

M の任意の二つの要素の対に対しては何れが他方に比してその性質をより多くもっているということを決定出来るものとする。

BはAより多くということを A→ B とする。

順位づけの本質はその移動律である。（A,B,C に対して A→B，B→C ならば必ず A→ C でなければならない）こういう性質がみたされないという要素に富んで居れば居る程順位づけが出来るという性質に欠けていることになる。

故に A→B，B→C 且 C→ A ; 又は B→A，C→ B 且つ A→ C なる関係をもつ A,B,C を廻転三対とよぶことにするとMに於て或る性質に関する廻転三対が少いということが，この性質が順位づけられるということに対しての一つのめやすになる。

n が奇数のときはこんな廻転三対の最大数は $\frac{1}{24}(n^3-n)$

n が偶数のときは 〃 〃 〃 $\frac{1}{24}(n^3-4n)$

であることが証明されているので廻転三対の数を d として，

$$\boxed{\begin{aligned} \varsigma &= 1 - \frac{24d}{n^3 - n} \qquad n = \text{奇数} \\ &= 1 - \frac{24d}{n^3 - 4n} \qquad n = \text{偶数} \end{aligned}} \cdots (31)$$

と定義し coefficient of concordance と呼ぶ。

$\varsigma = 1$ ならば $d = 0$ で data は順位をつけることが出来る。

ς の値（ d の値）の或る意味での有意検定は次の様にして出来る。即ち M の任意の二つの対に対して矢の向きが他のいかなる対に対しても独立にあらわれると仮定する。即ち2の ${}_nC_2$ 乗通りの矢じるしをもつた network がみんな等しい確率であらわれると仮定して d の分布を求める。n の大きさは

$$\nu = \frac{n(n-1)(n-2)}{(n-4)^2} \qquad \chi^2 = \frac{8}{n-4}\left\{\frac{1}{4}\binom{n}{3} - d + \frac{1}{2}\right\} + \nu \cdots\cdots (32)$$

とすると χ^2 は自由度 ν の χ^2 分布に漸近的に従うことが知られている。故に例えば或る観測された d の値 d_0 が χ^2 の 0.95％水準にある時 $d > d_0$ なる確率は $1 - 0.95 = 0.05$ である。

そういう時はそういう性質に関して順位をつけない方がよい。つまりこれは，順位づけすべきか否かに関して，或る意味で助言を與えるものである。

次に実際 d を勘定する方法を示す。又 M の要素に番号をうつてあるものとする。M の要素の名（番号）を縦横にならべた下の様な表をつくり，今しらべている性質で $j \to i$ であつたとすれば，$\delta_{ij} = 1$；$i \to j$ であれば $\delta_{ij} = 0$ 等と記録する。

そして $\sum_{j=1}^{n} \delta_{ij} = a_i$ とすると

$$\boxed{d = \frac{1}{12} n(n-1) - \frac{1}{2} \sum_{i=1}^{n} a_i^2} \cdots\cdots (33)$$

	1	2	3	…………	n	計
1	×	δ_{12}	δ_{13}		δ_{1n}	a_1
2	δ_{21}	×	δ_{23}		δ_{2n}	a_2
⋮						
n	δ_{n1}	δ_{n2}	δ_{n3}		×	a_n
計	b_1	b_2	b_3		b_n	

又 $\sum_{i=1}^{n} \delta_{ij} = b_j$ とすると

$\sum_{j=1}^{n} b_j^2 = \sum_{i=1}^{n} a_i^2$ なる関係があるから(33)に於てΣb_i^2

をつかつてもよい 又 検算にも用いられる。

§ XV Coefficient of agreement

次に上の表をもとにして m一順位づけと同じ様に数個のdata の一致性を考える。 今 m 人のうち l 人が $i \leftarrow j$ と判断した時は $\delta_{ij} = l$ 従つて $\delta_{ji} = m - l$ 等と記録する。

m 人が完全に一致しおれば δ_{ij} のうち ${}_nC_2$ は m と記録されている筈である。

$$\Sigma = \sum_{i \neq j} \binom{\delta_{ij}}{2} \quad \cdots\cdots\cdots\cdots \quad (34)$$

和は ${}_nC_2$ 個のすべての箱にわたる。そして

$$\boxed{\mu = \frac{2\Sigma}{{}_mC_2 \ {}_nC_2} - 1 = \frac{8\Sigma}{m(m-1)n(n-1)} - 1}$$

…………… (31)

と定義して μ を Coefficient of agreement と呼ぶ。

これは m 通りの判断が完全に一致しておれば +1 を与える。しかし最小値は m が偶数なら $-1/(m-1)$ であり m が奇数ならば，$-1/m$ である。それで最小値が −1 になるのは m＝2 の時に限る。

m＝2 の時は μ は τ と一致する故に μ はこんな意味での，τ の一般化と考えられる。

（31）の Σ は計算して

$$\Sigma = \sum_{i>j} s_{ij}^2 - m\sum_{i>j} s_{ij} + {}_mC_2\,{}_nC_2 \quad \cdots\cdots\cdots\cdots \quad (32)$$

$\sum_{i>j}$ は $\sum_{j>i}$ としてもよい　つまり表の X一 対角線の上半又は下半にわたって和をつくる。

M の対の中には「何が他方より」ということが判断出来ないものがあってもよい。そういう時例えば（i, j）について m 人中 l 人が $i \leftarrow j$　k 人が判断出来ず残りが $i \rightarrow j$ と判断したとすると

$s_{ij} = l + \frac{1}{2}k \qquad s_{ji} = m - l - \frac{1}{2}k$ と記録する。

Σ は（32）で計算すればよい。意味からすれば判断出来ないというのを判断出来るとして k 人の人について $i \rightarrow j$，$i \leftarrow j$ なる判断はすべて独立に同じ確立でおこるものとして（34）に於て平均をとるというやり方をすべきであるが k が小さければそうしたものと（32）で計算したものとあまりかわらない。

扨，もとにもどり network の $m \times {}_nC_2$ 個の矢の向きが at ramdom である時を考えると μ の分布が得られ又検出来る。

m, n が相当大きい時，即 $n > 7$（但し m＝3 の時は $n > 8$）位になれば近似的分布がつかえる。

$$\boxed{\begin{aligned} \chi^2 &= \frac{4}{m-2}\left\{\Sigma - \frac{1}{2}\binom{n}{2}\binom{m}{2}\frac{m-3}{m-2}\right\} \\ \nu &= \binom{n}{2}\frac{m(m-1)}{(m-2)^2} \end{aligned}} \quad \cdots\cdots\cdots\cdots \quad (33)$$

が自由度 ν の χ^2 分布 に漸近的に従うことが用いられる。

以上 Rank correlation methods の概要をのべたつもりだが冗長に過ぎた点，又書くべき点で見落したものもあることをお詫び申すと共に，粗雑な原稿を見て多くの誤りを指摘して下さった，畏友 鍋谷清治氏に心から感謝するものである。

附　　表

τ に対する S が特定の値に等しいか，それを超えるということの確率（Sの負の値に対しは全く対称）

S	n: 4	n: 5	n: 8	S	n: 6	n: 7
0	0.625	0.592	0.548	1	0.500	0.500
2	0.375	0.408	0.452	3	0.360	0.386
4	0.167	0.242	0.360	5	0.235	0.281
6	0.042	0.117	0.274	7	0.136	0.191
8		0.042	0.199	9	0.068	0.119
10		$0.0^{2}83$	0.138	11	0.028	0.068
12			0.089	13	$0.0^{2}83$	0.035
14			0.054	15	$0.0^{3}14$	0.015
16			0.031	17		$0.0^{2}54$
18			0.016	19		$0.0^{2}14$
20			$0.0^{2}71$	21		$0.0^{3}20$
22			$0.0^{2}28$			
24			$0.0^{3}87$			
26			$0.0^{3}19$			
28			$0.0^{4}25$			

（$0.0^{3}87$ 等は 0.00087 等の略）