㉔ 統計数理的数量化の問題補遺 *

(講究録第6巻第1.2.3.号参照)

林 知己夫

§0, Reliability と Validity 補遺

Reliability index と言はれているものを補つておこう。これは通常の本にもみられておるものである。

あるT.G.を固定しこれについてテストを行つたことを問題にする。T.G.の各個に等しい抽出確率を與へて母集としよう。

各人の点数は信頼度が全くあるとするとき $\overline{X}_1, \overline{X}_2, \cdots\cdots \overline{X}_N$ であるとする。これを総称して $\overline{X}$ とあらわす。

$$\begin{cases} \overline{X}_0 = \dfrac{1}{N}\sum_{i=1}^{N}\overline{X}_i \\ \sigma_{\overline{x}}^2 = \dfrac{1}{N}\sum_{i=1}^{N}(\overline{X}_i - \overline{X}_0)^2 \end{cases}$$

しかし各人の点数は Reliability なきため

$$X_1, X_2, \cdots\cdots X_N$$

と成る。これを総称して X とする。

こゝに

$$X_i = \overline{X}_i + e_i$$

とする。これを $X = \overline{X} + e$ と総称する。

e_i は random factor であり、Reliability をおとす因子である。

---

* 本研究は文部省科学試験研究費による研究の一部である。

$$E(e_i)=0 \qquad E(e_i^2)=\sigma^2$$

$$E(e_i \cdot e_j)=0 \quad (i \neq j), \quad E(\overline{X}e)=0$$

としておこう。 E は Reliability についての繰返し及び人についての expectation とする。

今 $\overline{X}$ と X との相関係数 r（母集団についてのもの）を以て Reliability の index と言ふことにする。

$$r = \frac{E(\overline{X}_i-\overline{X}_0)(X_i-\overline{X}_0)}{\sigma_{\overline{X}}\ \sigma_X} = \frac{\sigma_x}{\sqrt{\sigma_x^2+\sigma^2}}$$

$$\sigma_{\overline{X}}^2 = \sigma_x^2$$

$$\sigma_X^2 = \sigma_x^2 + \sigma^2$$

さて実際にはこれは求められない。 我々は第一回の調査第二回の調査の結果 $X(1)$ と $X(2)$ との関係しかわからない。

こゝに

$$X_i(1) = \overline{X}_0 + e_i(1),$$

$$X_i(2) = \overline{X}_0 + e_i(2)$$

但し $E(e_i(k))=0$

$E(e_i^2(k))=\sigma^2$

その他の条件は上述の通り。 $i \neq j$ ならば $X(1)\ X(2)$ は独立

この相関係数を ρ とする。（あらゆる場合について平均値をとつたもの，つまり母集団のものとする）

$$\rho = \frac{E\left(X_i(1)-\overline{X}_0\right)\left(X_i(2)-\overline{X}_0\right)}{\sigma_x^2}$$

$$= \frac{\sigma_x\sqrt{\sigma_x^2+\sigma^2}\ r}{\sigma_x^2+\sigma^2}$$

$$= \frac{\sigma_x}{\sqrt{\sigma_x^2+\sigma^2}}\,\gamma = \gamma^2$$

したがつて

$$r = \sqrt{\rho} \qquad となる。$$

つまり計算によるものの平方根として *index of Reliability* が考えられる。但しこれは母集団についての関係である。

しかも上述の條件あるものについてである。*Sample* については別様に考えねばならない。

§ 1　これは「第五章」*Scale Analysis* に関する個所の補遺である。66 P の 6.につづくものである。したがつて7.としておこう。

7.　*Scale Analysis* によつて得られた数 X は意義ある *Rank Order* をあたえるものである。$(X_i - X_j)$ は *Scale* の *distance* をあたえるものではない。

distance と言う様な *Metric* をつくるためには「*distance* なるものは『・・・に対して・・・』」と言ふ様なより高い見地からする評価的立場がとられねばならない。この *Scale Analysis* と言う内部操作丈によつてはつくれるものではない。

$X_i > X_j$ ならば i は j に較べて上(又は下)の位置にあると言うことを示してあるものにすぎない。

しかし前にものべた様に $X_i > X_j$ ならば各属性に於ても必ず上(下)位にあることを示しているものである事は注意すべき所である。

8.　この *Scale Analysis* は一次元的なものをつくり出す働きがあるのであるが，一次元的なもののみを問題にすればよいと言う様なことを物語っているのではない。

一次元的なもののみを問題にすれば，そのえらばれたものについて又は属性の *Reproducibility* はあるが「知りたい事」に対する *reproducibility* (*validity* と言える) はもっていな

いことになるのである。

validity と Reproducibility あるためには一次元的な分離グループを結合してゆかねばならない。

つまり「知りたい事」をもつともよくあらわすためには「如何に結合すべきか」の問題がおこつてくるのである。

これらについては私の考えを次に述べる。

9. ある母集団から選ばれたn人のサムプルの各人にR個の，item よりなる調査を行い，各 item の唯一つのカテゴリーにチエツクする様に指示するものとする。　又各人については，この調査とは別のものによつてある数量が得られているとする。

このとき，このある別の数量（Outside variable と言う。）と調査の item（カテゴリー）との関係をもとめ，Outside variable を最もよくあらわす様にするためには多くの item を綜合しいかに各カテゴリーに数量を與えるか。

こゝでのべた Scale Analysis によつて得られた Scale value をつかい Out side variable との関係を求めようとするこころみがあるが，これは全く Scale Analysis 及び Scale のいみの誤用である。

Scale Analysis は前述の様にこの様なことを目的としてつくられたものではない。

この問題は正に Scale Analysis の問題とは本質的に phase を異にしているのである。

もし Outside Variable が数量であらわされてゐるとするならばこれを最もよく表現し得る様な数量を各 item（その各カテゴリー）にあたえて，これを綜合すればよいのである。

つまり Outside variable のもつとも成功率の高い予測のために各人の behavour の patterning を綜合してゆくことを考えるのである。

各 item は Scale をつくつている必要は全然ないのである Scale のいみで次元をことにしていてもよいのである。　これを行う方法の一例を次に示してみよう。

今 outside variable をA；各 itemを$X_1, X_2, \cdots\cdots X_R$とする。

註： 前論文とは記号を異にするがこの方が計算に便のためこの方にかえるが此の点注意せられたい。

i なる人は variable で A_i をもち各 item で（$X_{1i}, X_{2i}, X_{3i} \cdots\cdots X_{Ri}$）なるカテゴリーの反応を示すものとする。

各 item のもつカテゴリー：これを順に示すと

$$\{C_{11}, C_{12}, \cdots C_{1k_1}\}, \{C_{21}, \cdots C_{2k_2}\}, \cdots \{C_{R1}, \cdots C_{Rk_R}\}$$

の通りになるが上述のもの X_{li} は l item のカテゴリー $\{C_{l1}, \cdots\cdots, C_{lk_l}\}$ の中の孰れか一つを示しているものである。

各 item に夫々 $k_1, k_2, \cdots k_R$ 個のカテゴリーがあるものとする。 この関係を図示してみよう。

item	X_1			X_2			----------	X_R		
outside variable \ カテゴリー	C_{11}	----	C_{1k_1}	C_{21}	----	C_{2k_2}	----------	C_{R1}		C_{Rk_R}
A_1		◍		◍					◍	
A_2	◍					◍		◍		
⋮										
A_i		◍			◍			◍		
⋮		X_{1i}			X_{2i}			X_{Ri}		

◍ 印は，各人（サムプル）の各 item で反応するカテゴリーを示す。

この様な場合，例えば（他の形でもよい！）

A と $X_1+X_2+X_3+\quad+X_R$ との相関係数を最大にする様に各 item のカテゴリーに数量をあたえることが考えられる。

これが一つの有効な数量化であることは言ふまでもないであろう。

今，総数（$k_1+\cdots\cdots+k_2$）個の各カテゴリーに数量 $x_{11}',\cdots\cdots x_{1k_1}$，$x_{21}\cdots\cdots x_{2k_2}$，$\cdots\cdots$，$x_{R1}\cdots\cdots x_{Rk_R}$ があたえられるものとする。

さて，各 item 各カテゴリーでの反応を示すものの数を，$n_{11},\cdots\cdots n_{1k_1},\cdots\cdots n_{R1},\cdots\cdots n_{Rk_R}$，であるとする。

調査されたランダム・サムプルの大きさを n とすると

$$n = \sum_{k=1}^{k_j} n_{jk} \qquad (j = 1, 2, \quad - R)$$

である。

今，$n_{ij} \geqq 2$ としておこう。

つまり各カテゴリーに属しているサムプルの数は2以上であるとする。

今 $x_{j(i)}$ は i なる人が j item で反応するカテゴリーをあらわす

$$x_{1(i)} + x_{2(i)} + \quad\cdots + x_{R(i)} = \alpha_i$$ とおく

そうして（A_i，α_i）の相関係数

$$\rho = \frac{\frac{1}{n}\sum_{i=1}^{n}(A_i-\overline{A})(\alpha_i-\overline{\alpha})}{\sigma_A\ \sigma_\alpha}$$

$$\overline{A} = \frac{1}{n}\sum_{i=1}^{n} A_i$$

$$\overline{\alpha} = \frac{1}{n}\sum_{i=1}^{n} \alpha_i$$

$$\sigma_A^2 = \frac{1}{n}\sum_{i=1}^{n}(A_i-\overline{A})^2$$

$$\sigma_\alpha^2 = \frac{1}{n}\sum_{i=1}^{n}(\alpha_i-\overline{\alpha})^2$$

を考えるのである。

ρ は，原点のとり方に依存しないので

$$\bar{A} = 0, \quad \bar{\alpha} = 0$$

（さらに各 item の平均を0となる様にしてもよい）

となるとしておこう。

こうすると

$$\rho = \frac{\frac{1}{n}\sum_{i=1}^{n} A_i \alpha_i}{\sigma_A \ \sigma_\alpha}$$

$$\sigma_A = \frac{1}{n}\sum_{i=1}^{n} A_i^2$$

$$\sigma_\alpha = \frac{1}{n}\sum_{i=1}^{n} \alpha_i^2$$

さて ρ が最大になる様に l item の m カテゴリーに数量 x_{lm} をあたえることにしよう。

このため

$$\frac{\partial \rho}{\partial x_{lm}} = 0 \qquad (l = 1, \cdots R, \ m = 1, \cdots k_{lm})$$

をつくる。

これを計算する前に次の様な記号をつくっておこう。

まづ $\delta_i(jk)$ をつくる。

$$\begin{cases} \delta_i(jk) = 1 & i \text{ なるものが } j \text{ item } k \text{ カテゴリーの反応を示す時} \\ \quad ,, \quad = 0 & \text{さうでない時} \end{cases}$$

こうすると

$$\sum_{k=1}^{k_j} \delta_i(jk) = 1$$

$$\begin{cases} \delta_i(jk)\delta_i(jk') = 0 & (k \neq k') \\ \qquad ,, \qquad = 1 & (k = k') \end{cases}$$

各人は各 itemでどれか一つの反応を示すことを現す。

$$\begin{cases} \sum_{i=1}^{n} \delta_i(jk) = n_{jk} \\ \sum_{k=1}^{k_j} \sum_{i=1}^{n} \delta_i(jk) = n \qquad \text{（すべての } j \text{ に対して）} \end{cases}$$

次に

$f_{lm}(jk)$ を定義する

$$\sum_{i=1}^{n} \delta_i(lm)\,\delta_i(jk) = f_{lm}(jk)$$

つまり（l item．mカテゴリー）の反応をしたものの中（j item　kカテゴリー）の反応を示すものの数である。

$$\sum_{m=1}^{k_l} f_{lm}(jk) = n_{jk} \qquad \text{（すべての } l, j \text{ に対して）}$$

$$\sum_{m=1}^{k_l} \sum_{k=1}^{k_j} f_{lm}(jk) = n \qquad \text{（すべての } l, j \text{ に対して）}$$

$$f_{lm}(jk) = 0 \qquad j = l,\ k \neq m$$

は明らかである。

さて

$$\frac{\partial \rho}{\partial x_{lm}} = 0 \quad \text{をつくると}$$

$$\frac{1}{n} \frac{\partial}{\partial x_{lm}} \sum A_i \alpha_i - \sigma_A \rho \frac{\partial \sigma_\alpha}{\partial x_{lm}} = 0 \quad *$$

$$\frac{1}{n} \frac{\partial}{\partial x_{lm}} \sum A_i \alpha_i = \frac{1}{n} \sum_{i=1}^{n} A_i \delta_i(lm)$$

$$\frac{\partial \sigma_\alpha}{\partial x_{lm}} = \frac{1}{n\sigma_\alpha} \sum_{i=1}^{n} \sum_{j=1}^{R} x_{j} \delta_i(lm)$$

$$= \frac{1}{n\sigma_\alpha} \sum_{i=1}^{n} \sum_{j=1}^{R} \sum_{k=1}^{k_j} x_{jk} \delta_i(lm)\,\delta_i(jk)$$

$$= \frac{1}{n\sigma_\alpha} \sum_{j=1}^{R} \sum_{k=1}^{k_j} x_{jk} f_{\ell m}(jk)$$

$$= \frac{1}{n\sigma_\alpha} \left(x_{\ell m} n_{\ell m} + \sum_{j=1}^{R}{}' \sum_{k=1}^{k_j}{}' x_{jk} f_{\ell m}(jk) \right)$$

$\Sigma'\Sigma'$ は，同時に $j=\ell$ $k=m$ となる場合をのぞくものとする。

したがって

※は，

$$\sum_{i=1}^{n} A_i \delta_i(\ell m) = \frac{\sigma_A \rho}{\sigma_\alpha} x_{\ell m} n_{\ell m} + \frac{\sigma_A \rho}{\sigma_\alpha} \Sigma'\Sigma' x_{jk} f_{\ell m}(jk)$$

$$(\ell = 1, 2, \cdots\cdots R, \quad m = 1, \cdots\cdots k_\ell)$$

となる。

$\bar{\alpha} = 0$，さらに又各 item での平均が 0 と言ふ條件は

$$\sum_m \sum_j \sum_k x_{jk} f_{\ell m}(jk) = \sum_j \sum_k x_{jk} n_{jk} = 0,$$

（すべての ℓ に対して）

又 $\displaystyle\sum_{k}^{k_j} x_{jk} n_{jk} = 0$ （すべての j に対して）

とかきあらわせる。

さて上の式をとけばよいのである

$\dfrac{\sigma_A \rho}{\alpha_x}$ はすべてに共通であるから常数とおき一次聯立方程式をとけばよいのである。

これを求めたのが $x_{\ell m}$ であるが，

今特別な場合を考えてみる。

$$\frac{\sigma_A \ell}{\sigma_\alpha} \Sigma'\Sigma' x_{jk} f_{\ell m}(jk)$$

が l を一定としたとき m の如何に関せず一定であるとする。

l item（lはすべてをうごく）と他の item とが統計的に独立であればこのことは母集団において当然みたされる。

さうすると條件により容易に

$$\frac{\sigma_A \rho}{\sigma_J} \Sigma' \Sigma' \ x_{jk} \ f_{lm}(jk) = 0$$

となることがわかる。

さうすると

$$x_{lm} = \frac{1}{n_{lm}} \sum_{i=1}^{n} A_i \delta_i(lm)$$

として x_{lm} は求められる。

この x_{lm} を第一次近似として用ふるのがよいであろう。つまり各 item が互に独立と假定したものを第一次近似として用ふのは自然であろうと思う。

（註）A_i と $X_{(i)} + \cdots + X_{R(i)}$ との相関関係ではなく A_i を $(X_{(i)}, \cdots\cdots, X_{R(i)})$ から linear な関係で推定しようとする場合は，量化された item，カテゴリーの値を用いて通常の様に重相関係数を求める考えから

$$\sum_{i=1}^{n} (A_i - l_1 X_{(i)} - \cdots - l_R X_{R(i)} - C)^2$$

を最小にすることを考えればよい。方法的には上記のものと全く同様である。

さて Outside Variable が数量的にあらわされている場合は behaviour の patterning の綜合は以上のことで解決がつくものと思う。

「讀み書き能力調査」で個人，地域の能力点数を推定する時に用いた考えはここでのべた所の第一次近似の数量化を用いたものである。実は，さらに計算すればもつと有効な数量化がある筈であった。

しかしこれ以上よく推定できると言ふことの保障があるから，information の損失はあるが計算の労を考えるとき一応の結果

であると言えよう。

Outside Variable が数量化されて居ない場合はこの Variable も，item，カテゴリーであらわされるのでこれに対する数量化の問題を更めて考えねばならなくなる。

この数量化の問題は前の論文，この論文の考えにしたがつて行はれればよいであろう。又 outside のものがカテゴリーであつても相関比を最大にする即ち全体の分散一定の下に各カテゴリー間の分散を最大にするように，数量化しさへすればよい，方法は全く同様である。

多少長くなつて Scale Analysis と論題ははなれたがこゝで計算し，論じた事は通常の社会心理調査問題でよく遭遇する問題であるから煩を厭はずのべてみた。

上述のことによつても解る様に Scale Analysis と Outside Variable の問題とは全く質を異にするのである。

Scale Analysis は全く内部操作であり，この内部操作によつて一次元的なもの（group）を見出すことにあるのである。

Out side のものの出現，これとの関係は Scale Analysis の問題ではない。

こゝではサムプルの n 人に対して数量化の方法を考えた。母集団における数量の推定は如何になるか。これは通常のサムプリングの考え方，即ち $y=f(x)$ の時の y の分散は

$$\sigma_y^2 = \left\{ f'(X) \right\}^2 \sigma_x^2$$

X は x の母集団の値，となる様な考え方にしたがつて與えた数量の信頼度は求められる。

§. 2. これは，第五章，第六章の次に入るべきもので新しい「第七章」とでもしたいところのものである。のべようとするところは第五章，第六章をつらぬくところの L. guttman の一貫した考え方であり，§ 1. 五章，六章，七章（第一，二，三号の第七章）で私の考えた所のもの（これも Optiman と言う立場で一貫している。）との相違を明白にすることである。

L. guttman は

(a) The Prediction of Personal Adjustment; A Survey of Logical Problem and Research Techniques, with Illustrative Application to Problem of Vocational Selection, School Success, Marriage, and Crime; Horst (P), Wallin (P), Guttman (L), Reed (R.B) Richardson (M.W)との共著; Social Science Research Council, New York, 1941; Bulletin 48

(b) Measurement and Prediction; Stouffer (S.A.) Guttman (L), Suchman (E.A.), Lazarsfeld (P.F.) Star (S.A.), Clausen (J.A.) との共著; Princeton University, 1950;

の中でのべている数量化(Scaleをあたえる方法)の考えとPaired Comparisonの考えはよく一致している。

(なお(a)の文献はきわめて興味深いものである)。

まづ,両書の考え方をまとめてのべておこう。

サムプルはn人でありitemはR個とする。各itemはk_i ($i=1, \cdots\cdots R$)のカテゴリーをもつとする。item R個の調査を各人に行って反応をみることにする。勿論各人は各itemで唯一つのカテゴリーをチェックするものとする。この時,例えば次頁の様な様相を得たとする。

この時全くアプリオリな假定をおかずに各item,カテゴリーならびに各個人に数量をあたえようとするのである。

こうすることが各個人のbehaviourを綜合的に予測するのに有用であると考えられるからである。数量をあたえると言っても勿論Scaleをつくると言ふ前提の下にすることである。

即ち,Scaleをつくるときのinternal Consistencyが保存される様に考えてするのである。

つまり;個人にあたえた数量と各itemカテゴリーにあたえた数量との関係が可能なるかぎり密接になる様に 即ち一方を知れば一方のことがなるべく正しく知り得られる様にするのである。

個人 \ item	X_1			X_2		・・・	X_R			
	C_{11}	C_{12}	C_{13}	C_{21}	C_{22}	・	C_{R1}	C_{R2}	C_{R3}	C_{R4}
1	V				V			V		
2		V			V		V			
3	V			V				V		
4			V		V				V	
⋮		⋮		⋮				⋮		
n		V		V			V			

こうすれば各個人の behaviour の pattern が一つの数量によつてある信頼度の下にあらわされ得ると言うことになつてくるのである。 これが Scale Analysis の考えに通ずる，さらに言えばその本質をつくものであることがすぐ了解せられよう。

つまり個人に與えられた一つの数量を知ることが各 item の反応（behaviour）の pattern を知ることになるからである。

このあたえ方を次にのべてゆこうと思う。 まず單純なところから始める。

まず

$$
\begin{array}{cc}
 & \text{人} \longrightarrow \\
\begin{array}{c}\text{item}\\ \text{カテゴリー}\\ \downarrow\end{array} & M_x = \begin{pmatrix}
x_1 & 0 & x_1 & 0 & \cdots & 0 \\
0 & x_2 & 0 & 0 & \cdots & x_2 \\
0 & 0 & 0 & x_3 & \cdots & 0 \\
\cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\
0 & 0 & x_4 & 0 & \cdots & x_4 \\
x_5 & x_5 & 0 & x_5 & \cdots & 0 \\
\cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\
0 & x_{s-3} & 0 & 0 & \cdots & x_{s-3} \\
x_{s-2} & 0 & x_{s-2} & 0 & \cdots & 0 \\
0 & 0 & 0 & x_{s-1} & \cdots & 0 \\
0 & 0 & 0 & 0 & \cdots & 0
\end{pmatrix}
\end{array}
$$

㊟　前の反応を，たて，よこ，ひっくりかえしたものを考える。

これは各人各反応をマトリックスに書いたもので，0，は反応しないこと $x_1, x_2, \cdots\cdots x_S$ は各 item カテゴリーにあたえらるべき点数とする。

$$S = \sum_{i=1}^{R} k_i \qquad \text{である。}$$

次に

$$\alpha_x = \frac{1}{Rn}\sum_{j=1}^{S} N_j x_j$$

α_x は M_x の 0 でないものの総平均である

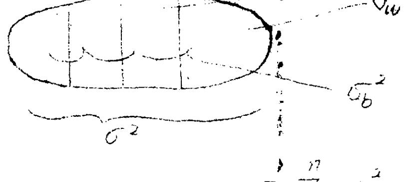
N_j は j カテゴリーに check する人の数である

この分散は

$$\frac{1}{Rn}\sum_{j=1}^{S} N_j x_j^2 - \alpha_x^2$$

次に α_i を M_x の i なる縦の R 個数量の平均とする。さうして相関比，即ち　全体の分散に対する縦の外分散の比，

（相関比　$\eta^2 = 1 - \dfrac{\sigma_w^2}{\sigma^2} = \dfrac{\sigma_b^2}{\sigma^2}$，　$\sigma^2, \sigma_w^2, \sigma_b^2$ は夫々全体の分散，内分散，外分散をあらわす。



に注意！）

$$\eta_x^2 = \frac{R\sum_{i=1}^{n} \alpha_i^2 - Rn\alpha_x^2}{\sum_{j=1}^{S} N_j x_j^2 - Rn\alpha_x^2}$$

を考えこれが最大になる様に x_j をきめるのである。　つまり全体の分散一定の下に各個人間の分散が最大になる様に各個人の中で item カテゴリーに関する分散が最小になる様に数量化する

ことを考えるのである。

又 η_x^2 は，原点のとり方によらぬから $\alpha_x = 0,\ \sum_{j=1}^{S} N_j x_j = 0$ とすると

$$\eta_x^2 = \frac{R \sum_{i=1}^{n} \alpha_i^2}{\sum_{j=1}^{S} N_j x_j^2} \qquad \text{となる。}$$

これが最大になる様に x_j をきめるのである。 これは微分法によつて求められる。 今まではitemカテゴリーに数量をあたえたが同様の考えで個人に数量をあたえるものとしよう。

これを $y_1, y_2, \cdots\cdots, y_n$ とすると反応のマトリツクスは

$$
\begin{array}{cc}
 & \text{人} \longrightarrow \\
M_y = \begin{array}{c} \\ \\ \text{item} \\ \text{カテゴリー} \\ \downarrow \end{array} &
\left(\begin{array}{ccccc}
y_1 & 0 & y_3 & 0 & \cdots\cdots & 0 \\
0 & y_2 & 0 & 0 & \cdots\cdots & y_n \\
0 & 0 & 0 & y_4 & \cdots\cdots & 0 \\
\cdots & \cdots & \cdot & \cdots & \cdots\cdots & \cdots \\
0 & 0 & y_3 & 0 & \cdots\cdots & y_n \\
y_1 & y_2 & 0 & y_4 & \cdots\cdots & 0 \\
\cdots & \vdots & \cdots & \cdots & \cdots\cdots & \cdots \\
0 & y_2 & 0 & 0 & \cdots\cdots & y_n \\
y_1 & 0 & y_3 & 0 & \cdots\cdots & 0 \\
0 & 0 & 0 & y_4 & \cdots\cdots & 0 \\
0 & 0 & 0 & 0 & \cdots\cdots & 0
\end{array}\right)
\end{array}
$$

ここで又前と同様に考えて

$$\eta_y^2 = \frac{\sum_{j=1}^{S} N_j b_j^2}{R \sum_{i=1}^{n} y_i^2}$$

但し　　b_j は item，カテゴリー j に反応したすべての
人の点数の平均である。

をつくり，これが最大になる様に y_i をきめる。

即ち各 item カテゴリーの間の点のひらきが最大になる様にきめるのである。

以上の二つの考え方によって x_j，y_i の数量化が行われるのである。実際にこれをもとめることは微分法，マトリックス理論の考え方によるのである。

なお解はその生ずべき η の値の最大なるものを求めるのである

これがその時の η_x，η_y となるのであり，この η_x，η_y を得べき x_j，y_i が求める数量となるのである（これについては後述する）

さてかうする操作はいかなる意味を持つであろうか。

数量化を Scale の意味で考えてみよう。つまり最初にのべた意味，個人にあたえる数量と item カテゴリーにあたえる数量とが深い関係をもつ様にする（こうなることが Scale 本来の意味）ことで考えてみよう。

今，各 item，カテゴリーに

w_1，　w_2，・・・・・・ w_S　の数量

各人に

Z_1，　Z_2，　・・・・・・ Z_n　の数量

をあたえるとしよう。

こうすると反応のマトリックスは次の様に書ける。

$$
M_{(w \cdot z)} = \begin{array}{c} \text{人} \longrightarrow \\ \begin{pmatrix}
(w_1\ Z_1) & 0 & (w_1\ Z_3) & 0 & \cdots\cdots & 0 \\
0 & (w_2, Z_2) & 0 & 0 & \cdots\cdots & (w_2\ Z_n) \\
0 & 0 & 0 & (w_3\ Z_4) & \cdots\cdots & 0 \\
\hline
0 & 0 & (w_4, Z_3) & 0 & \cdots\cdots & (w_4\ Z_n) \\
(w_5, Z_1) & (w_5, Z_2) & 0 & (w_5, Z_4) & \cdots\cdots & 0 \\
\hline
- & - - - & - - - & - - - & - - - & - - - \\
\hline
0 & (w_{s-3}, Z_2) & 0 & 0 & \cdots\cdots & (w_{n-3}, Z_n) \\
(w_{n-2}, Z_1) & 0 & (w_{n-2}, Z_3) & 0 & \cdots\cdots & 0 \\
0 & 0 & 0 & (w_{n-1}, Z_4) & \cdots\cdots & 0 \\
0 & 0 & 0 & 0 & \cdots\cdots & 0
\end{pmatrix} \end{array}
$$

item カテゴリー ↓

となる。　こゝでマトリックスの要素の相関係数が最大になる様に w 及び Z を求めようと言うのである。

「個人の数量」と「item カテゴリーの数量」との間の関係が密接な様にと言うことが，両数量間の相関係数が最大となる様にと言う言葉によつて統計数量的に表現せられているのである。（この表現は密接の一表現！）。

さてこの相関係数は原点のとり方によらぬから各数量の平均を 0 とするならば，相関係数

$$\rho = \frac{\Sigma\, wZ}{\sqrt{\Sigma\, w^2\, \Sigma\, Z^2}}$$

とかきあらわせる。

さて，この ρ が最大になる様に

$$\frac{\partial \rho}{\partial w} = 0, \qquad \frac{\partial \rho}{\partial Z} = 0$$

をつくり，w 及び Z のみたすべき聯立方程式をつくる。

（なお，此等の方程式にはすべての x y の値を常数とする様なもの，この時 $\rho=1$ なるものが解として含まれるがこれは現実的に意味をなさぬので解としてはあげぬことにする。ρ が1でない最大のものを求めるのである。）

こうすると，ここの w_j 及び Z_i の値が前の方法で求めた x_j 及び y_i の値と一致し，その時の ρ の値（$\rho=1$ を除く）が η_x，η_y の値と一致することが容易に知られるのである。（これは，w，Z，x，y に関する聯立方程式を書き下ろしてみれば $w(Z)$ のみたすべき方程式と $x(y)$ のみたすべき方程式とが一致することが直ちに知られる）

つまり，単独に x，y を求めるときの抽象的條件がここの ρ の値を最大にする（$\rho=1$ はのぞく）と言う具体的條件（Scale の本質を直接つく條件）と一致することは興味がある。

こう考えてくると，この意味で x，y を求める方法は Optimum な方法と言えるのである。

ここで，もし ρ の値が小であればそれらは Scale をつくらぬと言えるのである（item カテゴリーとサムプルとの関係に於て！）。

Optimum な方法でも密接な関係をつくり出し得ぬと言うことになるから一つの数量で全体の pattern を推し測ることが無理になってくるからである。

さらに，各 item が，すべて二つのカテゴリーしか含まぬ場合には，x_j y_i を求めるのに美しい数学的取扱が出來るのである。

方法論は全く同様である。

これは，Measurement and Prediction にのべられてあることの紹介である。

item は R 個，あるとする。

これらは完全な Scale をなしているとすれば，反応の型は次の通りになる。

反応型によってサムプルは次のR+1個の型にわかれる

	item カテゴリー									
	1	2	3	------	R	R+1	R+2	R+3	------	2R
0	1	1	1	------	1	0	0	0	------	0
1	0	1	1	------	1	1	0	0	------	0
2	0	0	1	------	1	1	1	0	------	0
⋮										
R−1	0	0	0	------	1	1	1	1	------	0
R	0	0	0	------	0	1	1	1	------	1

1 は item カテゴリーに反応したことを示す。

(1, R), (2, R+1) ------ (R, 2R) は R 個の item, カテゴリーを示す。 1〜R は Positive なカテゴリー, R−2R は negative な反応を示す。 したがつてどちらか一方へ反応すれば一方へは反応しないことになる。 されば可能なる型は上に示す R+1 通りあり, item カテゴリーが完全な Scale をなしておれば各人は R+1 のカテゴリーの孰れかにぞくすることになる。今完全な Scale がつくられて居るものとして反応型, 及び item カテゴリーに数量を与へる事を考へよう。

今 e_{ij} を定義する。

$$\begin{cases} e_{ij} = 1 & i \text{ なる型の人が } j \text{ なる item カテゴリーに反応する。} \\ \quad = 0 & \text{しからざるとき} \end{cases}$$

さうすると

$$e_{ij} = 1 \quad (1 \leqq j-i \leqq m) \qquad i = 1, \cdots\cdots, R$$

$$\quad = 0 \quad (\text{しからざるとき}) \qquad j = 1, 2, \cdots\ 2R$$

$e_{ij} = 1$ なら $e_{i,R+j} = 0$ である。

今 f_i を i なる型に居している人数とする。
n は総サムプル数とすれば

$$n = \sum_{i=0}^{R} f_i$$

となる。　次に F_j を j item カテゴリーに反応する人数とすると

$$F_j = \sum_{i=0}^{R} f_i e_{ij}$$

となる。

さて次に前と全く同様にして２Ｒ個の item カテゴリーに x_j なる点をあたえ，i なる型のもの（個人のグループ）に y_i なる数量をあたえるものとしよう。

こうして（x, y）の反応マトリックスにおける相関係数

$$\rho = \frac{\sum_{i=0}^{R}\sum_{j=1}^{2R}\left(y_i - \frac{1}{n}\sum_{n=0}^{R} y_n f_n\right)\left(x_j - \frac{1}{nR}\sum_{k=1}^{2R} x_k F_k\right) f_i e_{ij}}{\sqrt{\left[R\sum_{i=0}^{R}\left(y_i - \frac{1}{n}\sum_{n=0}^{R} y_n f_n\right)^2 f_i\right]\left[\sum_{j=1}^{2R}\left(x_j - \frac{1}{nR}\sum_{k=1}^{2R} x_k F_k\right)^2 F_j\right]}}$$

を最大にする様な x, y を求めればよい。　これをとくと（ρ は原点のとり方によらぬ。）

$$\begin{cases} \sum_{j=1}^{2R} x_j e_{ij} = \lambda R y_i & (i = 0, 1, \cdots\cdots R) \quad (1) \\ \sum_{i=0}^{R} y_i f_i e_{ij} = \mu x_j F_j & (j = 1, 2, \cdots\cdots 2R) \quad (2) \\ \lambda\mu = \rho^2 & \end{cases}$$

となる。

これを満足する様な（x_j y_j）を求めればよいのである。
（第二式の x_j を第一式に代入すれば y のみの式となる）

これをとくのに定差方程式の考えを用うる。
(1)を i について第一次階差をとる。　これに (2) を代入して計算し，これにつゞいて第二階差をとる。これを変形すると結局

$$\begin{cases} C_i = F_{i+1}F_{i+R+1} & (i=0,1,\cdots\cdots R-1) \\ \phi \doteq \dfrac{n}{R\lambda\mu} \end{cases}$$

とおけば

$$\sum_{h=0}^{R} y_h C_{hi} = \phi f_i y_i \qquad (i=0,1,\cdots R) \qquad (3)$$

を得る。　ここに C_{hi} は次のマトリックスの element である
$(h, i = 0, 1, \cdots\cdots R)$

$$\begin{pmatrix} C_0, & -C_0 & & & & & \\ -C_0, & C_0+C_1, & -C_1, & & & \text{\Large 0} & \\ & -C_1, & C_1+C_2, & -C_2, & & & \\ & & -C_2, & C_2+C_3, & -C_3 & & \\ & & & \cdots\cdots & \cdots\cdots & & \\ & \text{\Large 0} & & & -C_{R-2}, & C_{R-2}+C_{R-1}, & -C_{R-1} \\ & & & & & -C_{R-1}, & C_{R-1}. \end{pmatrix}$$

x_j についても同様なものを得る。

以上の方程式をとくことになる。　これをといてもとまる ϕ の＝中の最大となるものが我々のもとめる ρ となるのである。

勿論，$\rho=1$ なるもの，即ち各 x, y が Const となるものが解として含まれるがこれはのぞくことにする。（現実的に意味をなさぬ）。

我々としてはその様に大なる ρ を求めるのである。

さて，上述の方程式をとくとき定差方程式の性質により，

(i)　(3)をみたす ϕ は $R+1$ 個あり，それはみなさとなる.

(ii)　$\rho=1$ でない (3)をみたす ϕ の値の和は 1 である.

(iii)　解 y_i　$(i=0\ \text{----}\ R)$ の $R+1$ 個の system には符号

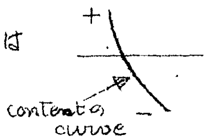
変化のないもの，一つの符号変化のあるもの，二つの符号変化のあるもの，……R個の符号変化のあるものが，存在する。しかもそれに応ずる ρ の値が大なるものほど符号変化が少い。

即ち $\rho_0=1,\ \rho_1>\rho_2>\cdots\cdots>\rho_R$ とすれば

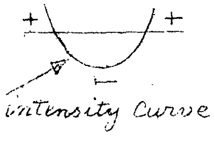
ρ_0 に応ずる y_i の Systemは符号変化なく
ρ_1 〃 〃 一つの符号変化あり
ρ_2 〃 〃 二つの 〃

なることがわかる。

Guttman はこれを解釈するのに ρ_1 に応ずる y_i の system は Scale Analysis の合計点に相当する Content の数量（これは



となる。つまり高いものから低いものに至る）である。ρ_2 に応ずる y_i の system は Intensity Analysis の intensity の数量に相当する intensity の点（これは



の型をなすから，この解釈は許される。つまり両極に強く中間に弱い！）をあたえるとのべてある。ρ_1 の方はきわめてよく了解されるのであるが，ρ_2 の方は私にはわからない。甚だ巧妙にはみえるが唯それ丈のことであつて何故に Intensity と解するかの論理的な必然性はない様に思われるからである。この点さらに研討の要があるものと思われる。

なお前論文でのべた各問に 4, 3, 2, 1, 0 の点をあたえ総得点を以て個人の点数としたが，これは一つの簡便な y_i のあたえ方であると考えてよい。（たとえ理論的には妥当でなくともあるいみで一応対応のつくものと見てよい。このことも前参照）

以上によって Scale Analysis の統計数理的基礎的な考え

を紹介したのであるがこの考え方は Guttman 流の quantification の底をながれているものと思われる。

第六章でのべた Paired Comparison の考え方正に然りである。それは全体の分散一定の下に between の分散を最大にすることであった。

これは $\dfrac{\sigma_b^2}{\sigma^2}$ と書けるのであつて、ここでのべた相関比 η^2 を最大にすることと全く等しいものなのである。

「相関比を最大にすること」即ち「全体の分散一定の下に between の分散を最大にすること」これが L. Guttman 流の根本思想であると思う。しかし現実的意味づけに於ては Scale Analysis と paired Comparison の時とはことなつてゐる。Scale Analysis の場合さうする操作は具体的意味をもち、Optimum なものと考へられたが、paired Comparison ではそれによる数量化は Optimum なものとは考えられないのである。（これについて前の論文参照）

ここの Guttman 流の考えは定性的な性質をもつものを分別する時にも用いられている。

例えば Johnson (P.O) The Quantification of Qualitative Data in Discriminant Analysis (Journal of American Statistical Association 1950)

これにはいくつかの Item（カテゴリー）をもつものを数量化しことなる Item（カテゴリー）をもつ各人がなるべくことなる様に（分別できる様に）するための数量化が考えられている。

これには数量をあたえ全体の分散（各人及び item カテゴリーに関するもの）一定としたとき各人間の分散が最大になる様に数量化すことが考えられている。

或は又 A group B group に属するいく人かの人がありこれがいくつかの item（カテゴリー）である反応を示すものとする。このとき item（カテゴリー）を数量化するのにそこで与えられる数量は A, B 各 group に属するものをなるべく、

それと解らせる様にすることを考える。

つまり *item* の反応による類型化のための問題とも考えられるのである。

このとき数量は，全体の分散一定の下にAB間の *between* の分散を最大にすることが考えられる。（*group* はいくつあつても同様）　しかし，これは *Optimum* なものであらうか。私にさうは思えない。

分別とは如何なる意味に於てか（分別の知識は何に用いられるか）を明白にし，この立場から有効な方法が考へらねばならない。と思ふ。　この意味で前論文の第七章の分別の考えは明瞭なものをもつているものと思ふ。

内部操作は一つの立場ではあろが，その意味が現実的に明白ではない場もあることに注意すべきである。

（追補記事末尾にあり）

§ 3.　第七章の末尾「学校テスト」と *follow up* の問題に関係する所である。

入学試験の点数と，その生徒の入学後の成績とを調査し，入学試験の合格点数をどうきめれば将来の予測成功率

$$\frac{\text{その中成績のよかつたもの}}{\text{ある合格点以上のもの}}$$

が最大になるであろうか。

このきめ方を合理的にする問題である。　この問題に限らず，この種問題に利用出来ることもあるのである。

これは，*The Prediction of Personal Adjustment* の *P. Horst* の考えである。

i なる生徒の $\left\{\begin{array}{l}\text{入学試験の点数を } x_i \\ \text{入学後の成績を } y_i\end{array}\right\}$ とおき $(x\ y)$ の同時分布 $f(x,y)$ を考える。

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)\,dx\,dy = 1 \quad \text{である。}$$

「成績のよかつたもの」と言うのを入学後の成績の y_0 点以上のものとしよう。 x_0 を入学の合格点数ときめる。

この時 x_0 以上であり y_0 以上であつたものは

$$\int_{y_0}^{\infty}\int_{x_0}^{\infty} f(x\,y)\,dx\,dy$$

である。

x_0 以上のものは

$$\int_{-\infty}^{\infty}\int_{x_0}^{\infty} f(x\,y)\,dx\,dy$$

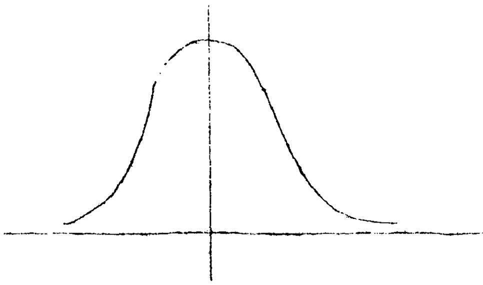
である。

したがつて

$$\frac{\int_{y_0}^{\infty}\int_{x_0}^{\infty} f(x\,y)\,dx\,dy}{\int_{-\infty}^{\infty}\int_{x_0}^{\infty} f(x\,y)\,dx\,dy}$$

を x_0 について最大になる様にすればよいのである。

§ 4. これは第9章としてもよい。 よく昔の方法では根本にある潜在的構造の假定をもとに置いて論議してゐるものがある。 この假定は機能的なものではないものがある。



例えはある特定の個人にある刺戟，R_a，をあたえたときそれに応ずるプロセス S_a は $\bar{S_a}$，分散 σ_a^2 の正規分布をなすものとして假定せられる．等のことである。（常に $R_a \to S_a$ と一意的に対応せず分布をもつと考えるのである）．この立場に立つて S_a 云々を比較し種々の事を論じようと言ふのがある。 例えば二つの刺戟 R_a, R_b の辨別の問題である．

S_a, S_b は弁別に参與し、判断を規定する二つのプロセスであるとする。 二つの刺戟をあたえたとき $S_a > S_b$ であるとするならば「R_a は R_b より強い（この言葉はなんでもよい）」と判断すると解釈するのである。

なお、S_a と S_b は統計的に独立であるとする。

$$S_a - S_b = S \qquad \text{とおくと}$$

$$E(S) = \bar{S} = \bar{S}_a - \bar{S}_b$$

$$E(S - \bar{S})^2 = \sigma_S^2 = \sigma_a^2 + \sigma_b^2$$

今ランダムな n 回の試行中 n_1 回 $S > 0$ となり $R_a > R_b$ と判断されたとする。

この時 $\bar{S}$ はどの位であると判定したらよいかと言ふ問題が考えられる。 この $\bar{S}$ の大さで各任意のいろいろな刺戟の差を比較しようと言うのである。 これは *Paired Comparison* の基本的問題である。

つまり二つのものを比較して " *prefer to* " の関係を見出し、しゅじゅの *pair* の比較から各刺戟の *Rank Order* をさぐらうと言うことになるからである。

数式的に書くと

$$\frac{1}{\sqrt{2\pi}\,\sigma_S}\int_0^{\infty} e^{-\frac{(S-\bar{S})^2}{2\sigma_S^2}}\,dS = \frac{n_1}{n}$$

$$\frac{1}{\sqrt{2\pi}}\int_{-\frac{\bar{S}}{\sigma_S}}^{\infty} e^{-\frac{t^2}{2}}\,dt = \frac{n_1}{n}$$

σ_S は常に一定と考え $\sigma_S = 1$ と仮定するものとする。このとき

$$\frac{1}{\sqrt{2\pi}}\int_{-\bar{S}}^{\infty} e^{-\frac{t^2}{2}}\,dt = \frac{n_1}{n} = p$$

$\left(\frac{n_1}{n}\right)$ を知ることにより数表を用いて $\bar{S}$ を推定し得られる。

この推定 $\bar{S}'$ の分散はどうであろうか

$$p = f(\bar{S}) \qquad \text{とすれば}$$

$$E(dp)^2 = \{f'(\bar{S}_0)\}^2 E(d\bar{S}')^2$$

ここに $\bar{S}_0$ は p の母集団のあたいに対する $\bar{S}$ の値である。

$$\sigma^2_{\frac{n_1}{n}} = \{f'(\bar{S}_0)\}^2 \sigma^2_{\bar{S}'}$$

故に現実的には

$$\sigma^2_{\bar{S}'} = \sigma^2_{\frac{n_1}{n}} \Big/ \left\{\frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}\bar{S}'^2}\right\}^2$$

$$= \frac{p(1-p)}{n} \Big/ \left\{\frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}\bar{S}'^2}\right\}^2$$

と考えてよい。

多くの場合この種のものでは「ランダムな n 回の試行」を「ある個人よりなる母集団からのランダムサムプル」と考えて調査を行い結論を出している様であるが，これがみとめられるためには，「個人の S_a の分布が，母集団の各個について一回の試行を行ったときそのときの分布と一致する」つまり「個人の多数回試行結果の分布が母集団各人の一回の試行結果の示す分布と一致する」即ち「時間的分布が空間的分布と一致する」と言うことが必要となってくるのである。よほどうまい母集団をとって来ないかぎりこれはみとめられないのは明らかである。したがってこの調査の考え方には少なからぬ問題が残るのである。しかも超えがたい困難さであろう。

さらに又正規性の假定も機能的なものでなく理論の本質が全くその假定によって居るのである事に思いを致せばますます好ましくないものとなって来よう。

この様に潜在的な Process に假定をおく方法の Validity は実例を以てその有用性がたしかめられざるかぎり non-sense なものと言うことが出来る。　（追補記事末尾にあり）

計算等に便利なこの方法もさう発展するものとは思へない。

§5. 第八章の文献の補遺として最近

Kuhn }
Tucker } 篇 *Contribution to the Theory of games*, Princeton 1950

がある様である。内容は数学的なものであり，條件の精密化，細部の数学的発展がのべられてゐる。

§6. Sociometry における一つの数量化。

Sociometry（この考へ方については最近 Moreno (J.L), Sociometry, Experimental Method and the Science of Society, Beacon House, 1951, がある）における各個人間の interaction を表現する方法の一つとして matrix 表現がある。

これを数量化する一つの試みについて考へたことをいささかのべてみよう。

［interaction の一つの表現］

人	1.	2,	3,	4,	…………	j	n
1	e_{11}	e_{12}	e_{13}	e_{14}		⋮	e_{1n}
2	e_{21}	e_{22}	e_{23}	e_{23}		⋮	e_{2n}
3						⋮	⋮
⋮						⋮	⋮
i	…………	…………	…………	…………	…………	e_{ij}	⋮
⋮							⋮
n	e_{n1}	e_{n2}	e_{n3}	e_{n4}	…………	…………	e_{nn}

e_{12} は1なる人の2なる人への態度，e_{21} は2なる人の1なる人への態度，即ち，1なる人が2なる人からいかに思はれてゐるかといふこと，e_{ij} は i なる人が j なる人に対する態度，

e_{ji} は j なる人の i なる人への態度即ち i なる人が j なる人からどう思はれてゐるかと言ふこと，したがつて i row は i なる人の他の人に対する態度，j Column は j なる人が他の人からどう思はれてゐるかと言ふ事をあらはすものとする。

この e_{ij} はあらかじめ Valid ある方法で数量化されて居るものと考へよう。

一般に $e_{ij} \neq e_{ji}$ は明らかである。

この時，相互の interaction をあらはしてゐる Matrix 表現をいかにまとめあげて一つの集団構造即ち各人がいかなるグループをつくつてゐるかを見たいものとしよう。

つまり親近性あるものは近く，反撥するものは遠く，グループされるのをみたいものとしよう。

今かりに e_{ij} は i から j への親近性の度合をあらはし，大なる数値をもつものはより親近性があるものと考へておく。

このため i なる人に x_i なる数量をあたへることを考へよう。

この x_i の値の様子によつてグループされる模様をみようとするのである。x は親近性をみるための一つのインデックスであると言へよう。さてそれならば，いかにして與へたらよいであらうか

$$Q = \sum_{i=1}^{n}\sum_{j=1}^{n} e_{ij}(x_i - x_j)^2$$

を x_i のもつ分散 $\frac{1}{n}\sum_{i=1}^{n} x_i^2 - \bar{x}^2$; $\left(\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i\right)$
一定の下に最小にするように與へてみよう。つまり e_{ij} の大なるものは x_i と x_j とがちかく e_{ij} の小なるものは x_i と x_j とが遠くなるようにあたへるのである。

従ってこれは一応合理的な立場であると言へよう，

この時式の性質から平均 $\bar{x}$ はどこにあつてもよいから $\bar{x} = 0$ と考へ

$$f = Q + \lambda' \left(\frac{1}{n} \sum_{i=1}^{n} x_i^2 - a^2 \right),$$

a, λ' はある常数

を x_i につき最小にするようにすればよい。

$$\frac{\partial f}{\partial x_i} = 0 \qquad (i = 1, 2, \cdots n)$$

を求めると

$$\left\{ \sum_{j=1}^{n} (e_{ij} + e_{ji}) \right\} x_i - \sum_{j=1}^{n} (e_{ij} + e_{ji}) x_j = \lambda x_i$$

$$\left\{ \sum_{\substack{j=1 \\ i \neq j}}^{n} (e_{ij} + e_{ji}) \right\} x_i - \sum_{\substack{j=1 \\ i \neq j}}^{n} (e_{ij} + e_{ji}) x_j = \lambda x_i$$

但し $\lambda = -\dfrac{\lambda'}{n}$

今 $e_{ij} + e_{ji} = a_{ij}$ とおくと

$$\left\{ \sum_{\substack{j=1 \\ i \neq j}}^{n} a_{ij} \right\} x_i - \sum_{\substack{j=1 \\ i \neq j}}^{n} a_{ij} x_j = \lambda x_i \qquad (i = 1, 2, \cdots n)$$

この a_{ij} のマトリックスは明らかに対称である。
したがつて逐次近似法の考へによつて，上式から λ（$\lambda = 0$ 以外のもの）及びそれに応ずる x_i を求め（x_i は比のみきまる），これから $\frac{1}{n} \sum_{i=1}^{n} x_i^2 = a$ によつてさいごの x_i がきまつてくる。

この x_i のかたまり工合によつてグループの型（一つの構造）が知られるのである。

今，マトリックスを

人	1	2	3	4
1	//////	2	-2	-1
2	2	//////	-1	-1
3	-2	-1	//////	1
4	-2	-2	2	//////

とすると，方程式右辺のマトリックスは

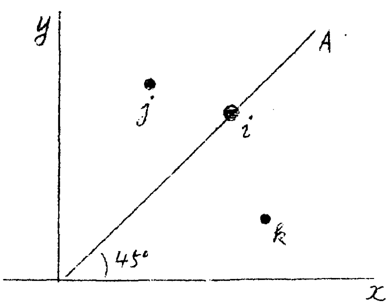
$$\begin{pmatrix} -3 & -4 & 4 & 3 \\ -4 & -1 & 2 & 3 \\ 4 & 2 & -3 & -3 \\ 3 & 3 & -3 & -3 \end{pmatrix}$$

となる、これから x_i をもとめると

$$x_1 = 1 \qquad x_2 = 0.75, \qquad x_3 = -0.89 \qquad x_4 = -0.87$$

となる.

しかしこれでゆくと a_{ij} のみがきいてくるので工合のわるいこともある。したがってこのようなときには二次元的に考へ「他の人に対する態度」を一つの次元，「他人からうけとるもの」を一つの次元と考へ，各人を二次元の上の点と考へようとするのである。このため i なる人に（x_i，y_i）なる点をあたへるのである。



x，y 軸を直交座標とするとき，x，y 軸の間の45°の線 A 上の点は「他の人への態度」と「他人からうけとるもの」とが同じ様な傾向にある人をあらはしてゐる。

このAを外れるものも人間関係では多いことであらう。一方のものがあるものに親近性をもつが，その人からは反撥をうけてゐると言ふ様な場合がこれである。（x_i，y_i）をあたへるのに前の場合と同様な考への下に

$$\sum_{i \neq j}\sum \left(\sum_k \frac{1}{(e_{ik} - e_{jk})^2} \right) (x_i - x_j)^2 + \lambda \left(\frac{1}{n} \sum x_i^2 - a^2 \right)$$

$$\sum_{i \neq j}\sum \left(\sum_k \frac{1}{(e_{ik} - e_{jk})^2} \right)(y_i - y_j) + \lambda'' \left(\frac{1}{n} \sum y_i^2 - b^2 \right)$$

を最小にするように (x_i, y_i) をきめればよい。

計算は

$$\sum_k \left| \frac{1}{(e_{ik} - e_{jk})^2} \right)$$

等をあらためて e_{ij} と考へれば全く同様にゆく。

なほ， $\sum_k (e_{ik} - e_{jk})^2 = 0$ ならば常に $x_i = x_j$ とるものとする。

さて

$$\sum_k \left(\frac{1}{(e_{ik} - e_{jk})^2} \right)$$

の意味をみよう。

$$\left\{ \begin{array}{l} x_i \cdots\cdots e_{i1},\ e_{i2}, \cdots e_{in} \\ x_j \cdots\cdots e_{j1},\ e_{j2}, \cdots e_{jn} \end{array} \right\}$$

i と j との他の人へ対する態度のくひちがひの程度は一応

$$\sum_k (e_{ik} - e_{jk})^2$$

によつてあたへるであらう。

したがつて

$$\sum_k \frac{1}{(e_{ik} - e_{jk})^2}$$

はくひちかひの大なるほど小，くひちがひの小なる程大となる傾向がみられるであらう。 したがつて

$$\sum_{i \neq j}\sum \sum_k \left(\frac{1}{(e_{ik} - e_{ij})^2} \right)(x_i - x_j)^2$$

を小にすると言ふ事はくひちがひの小なるほど左の項が大になるので x_i と x_j とを近くしなければならないと言ふ事を要求してゐることになるのである。 つまり他人への態度が似てゐるほど

x_i と x_j とが近くならねばならぬと言ふ傾向にしようとしてゐることなのである。 これがグルーピングするときのねらひであるのは言ふまでもない。

「他人からうけるもの」に対してもそれが近いものは近くあつまると言ふ様なことにするのであるから全く同様な論法でゆけるのである。

以上は論をすすめる中に多少假定が入るが一応は合理的立場のものであらうと思はれる。 しかしこゝの様に２乗をとつたり，逆数をとつたりすることをアプリオリに行はず，ある立場（Validity 或るいは Reproducibility）の徹底した立場から interaction patternings を数量化する考へに則ることがさらに望ましいのは言ふまでもない。 ここでの論述は一つのモデルとして Sociometric Representation をまとめる可能性を示したものにすぎない。

§ 7 附録 Non-parametric Test に関する文献抄

主な理論的文献は Wilks の Order Statistics にくわしいのでそれを参照せられたい。

手軽につかえるものとして

Olmstead (P.S) and Tuckey (J.W)

A Corner Test for Association, Annals of Mathematical Statistics 1947 を追加しておこう。

§ 8 補　前論文の主なミスプリントの訂正及び補

頁	行	誤	正
3	14	Who shall survive? 1949 のあとにつけ加える.	(Metal and Nervous Desease Publishing Comp.)
7	最終行	$r'_{nr} =$	$r =$
8	1	統計学	統計数理
8	9	一つの指標	一つの係数
8	下から3行目	(但し X_i, Y_i は ‥‥)	(但し $\{X_i, Y_i\}$ は i に関して ‥‥)
11	下から5行目	を考えることにしよう. あとにつけ加える	E は pattern と n 問題の現れ方に関する expectation をとることをあらわす
13	下から3行目	‥‥存在する程度をも‥	存在する程度，問題の質の異なる程度をも ‥‥
14	3行目 1図	問題は1から10まであり，その累積頻度は ‥‥‥ 図	問題は1から10まであり，i 番目の出来るものは必ず j 番目の問題も必ず出来る，j 番目の出来ぬものは必ず i 番目のものは出来ぬ ($i<j$ なる関係あるものとする) ものとする。

頁	行	誤	正
			この時成績の累積頻度が
14	下から12行目	難易が	難易及び重要性が
16	4行目	述べた様な立場から	述べた様な実証的立場から
65	最終行	一番最後ニツケ加エル	（但し (イ)(ロ)(ハ)(ニ)(ホ) になるに従って得点は下って居らねばならない）
80	7行目	アトニツケ加エル	しかしそうも云えない場合もある。これは被調査者のグループの性質によるものである。被調査者の意見の比較はこの *neutral point* の比較によって行い，解釈することも出来るのである。
92	下から2行目	$\Sigma\Sigma x_{kr} h_{jk/pr} = \frac{1}{2}E^2(x_{jp} + \Sigma\Sigma \cdots)$	$\Sigma\Sigma x_{kr} h_{jk/pr} \equiv \frac{1}{2}E^2(x_{jp} A_{jp} + \Sigma\Sigma \cdots)$
104	後 3行	$P = \frac{1}{\sqrt{2\pi}} \int_{\frac{m'+m}{\sigma'+\sigma}}^{\infty} \cdots$	$P = \frac{1}{\sqrt{2\pi}} \int_{\frac{m'-m}{\sigma'+\sigma}}^{\infty}$
110	3 行	$\frac{a_1(m_1+m_1') \pm \cdots}{2\sqrt{}}$	$\frac{a_1(m_1-m_1') + \cdots}{2\sqrt{}}$

頁	行	誤	正
122	後から2行	$\sum_{i}^{\beta_1}\sum_{j}^{\beta_2} d_{ij}\xi_i\xi_j =$	$\sum_{i=1}^{\beta_1}\sum_{j=1}^{\beta_2} \alpha_{ij}\xi_i\eta_j =$
140	21	5. H.B. Mam ···	追加 (Econometrica 1945)

Supplements to "The Quantification of Qualitative Data from the Mathematico Statistical Point of View."

Chikio Hayashi.

(See "The Quantification of Quantilative Data from the Mathematico-Statistical Point of View" Vol 6, 1.2, B)

This papers supplies the previous one.

(i) The quantification problem that requires the relation between on outside (riterion) variable and test items by quantifying the categrories of test items is solvet, from the Stand point of raing the efficincy of prediction. It is different from Scale Analysis. This point is discussed in detail.

(ii) mathematical construction of scale Analysis is discussed under the idea of L. Guttman. This idea leads to his theory of paired comparison and the difference between the idea of L. Guttman and mine is shown.

(iii) The problem of latent structure for example the structure of paired comparison is considered.

追補記事

なほ F. Mosteller は A theory of Scalogram Analysis, Using non-cumulative Type of Item, A new Approach to Thurstone's Method of Scaling Attitude (Laboratory of Social Relations, Harvard Univ, Report No. 9.) において Thurtone の Attitude Measurement と Guttman の Scale Analysis の考へとを形式の上で綜合しようとする試みをのべてゐる。しかしその統計数理的操作をみるとき全く Scale Analysis における数量化の数式を Thurstone 流のものに形式的に適用したものにすぎない。

従つて，これによつてあたへられる Thurstone Scale の数量は一体何を現してゐるのであらうか。このReproducibility は Thurstone の何をあらはしてゐるであらうか。全く疑問である。Thurstone の眞のねらひはこの数量化の方法では達せられないし，又，Thurstone 式意見のこの方法による数量化では Guttman の意味の Reproducibility もきはめてひくいことであらう。(これらについては前の本論文参照)

F. Mosteller の試みはこの両者の本質を解せぬものと言はねばならない，

同巧のものであるが，F. Mosteller は Some Miscellaneous Contribution to Scale Theory. Remarks on the Method of Paired Comparisons (Laboratory of Social Relations, Harward Univ. Report No. 10) において くの paired Comparison の結果をまとめる一つの方法を出してゐる。これによると S_i, S_j が感覚によつて知覚せられる確率変数とする。

平均は $\bar{S}_i$, $\bar{S}_j$ とする，

又 $E(S_i-\bar{S}_i)^2=\sigma^2$ (i によらぬ)

S_i と S_j との相関係数

$$\rho_{S_iS_j}=\rho_{ij}=\rho \quad (\text{一定})$$

とする。

今，$S_i - S_j = T_{ij}$ とおくとき T_{ij} は正規分布をなし

$$p_{ij} = P_r\{S_i > S_j\} = \frac{1}{\sqrt{2\pi}\,\sigma_{ij}'}\int_0^\infty e^{-\frac{\{T_{ij}-(\bar{S}_i-\bar{S}_j)\}^2}{2\sigma_{ij}'^2}}\,dT_{ij}$$

但し $\sigma_{ij}^2 = 2\sigma^2(1-\rho^2)$

が成立してゐるものとする。

簡單のため $\sigma_{ij}^2 = 1$ としておく，さうすると

$$p_{ij} = \frac{1}{\sqrt{2\pi}}\int_{-(\bar{S}_i-\bar{S}_j)}^{\infty} e^{-\frac{t^2}{2}}\,dt$$

となる。

今 p_{ij} は調査によって p_{ij}' と推定されれば

$(\bar{S}_i - \bar{S}_j)$ は $\bar{S}_{ij}'$ として推定される。

したがつて調査と眞に知らうとするものとの差

$$\sum_i \sum_j (\bar{S}_{ij}' - (\bar{S}_i - \bar{S}_j))^2,$$

つまりすべてのものについての誤差の和が最小になる様に $\bar{S}_i$, $\bar{S}_j$ を推定すると言ふ考へ方が用ひられる。

これも一つの方法であらうと思はれるが，潜在的構造の想定は面白くない点である。

§8. 後書き

以上統計的数量化の理論的なことについてのべてきたのであるが，こゝで数量化の立場をまとめておかう。

数量化は知らうと欲することを正に知り得られるように行はねばならない。このために Validity（妥当性），Reliability（信頼性），Objectivity（客観性），Reproducibility

（再現性），Consistensy（無矛盾性），Adequacy（適応性）ある有効なものでなくてはならない。（これらの性質については，青山博次郎，松下嘉米男，林知己夫，水野坦，社会現象の統計数理，朝倉書店，1951参照）

数量化は具体的な，現実的目標，現実の行動の有用な指針となるような目標をはっきりするどく定め，これをよく知り得られるように，種々の測定の結果にもとづいて，それに対応づけて行はれねばならない。 多くの場合効率の高い数量化を行ひ得たのは具体的目標が鋭く，明確な場合であつた。

数量化はこの意味で絶対的なものでなく，「知らうとすること」によつてかはりうる，機能的，操作的なものである。 同じものでもそれぞれの目標によつていろいろことなつて数量化されることになるのである（このときの立場は精度よく，性能よく，信頼度高く「知らうとすること」を如実に知り得られるようなインデックスをつくると言ふことである）

さて数量を與へるときの立場をまとめてみると，

(1) 現象に内容的構造的假設を置いて数量化するもの，
例へば，paired Comparison（第九章）における潜在的構造（正規分布）等である。

(2) 假設をおかぬもの

(イ) Outside の Criterion あるもの
外的な規準がありこれをもつともよく表現する様に behavior の patterningsを数量化するもの，
こゝでは Validity と言ふものが大切なものとなつてくる，これは現象豫測の問題，回帰面的推定の数量化等

(ロ) Outside Criterion ない場合
このときは各 behaviour patterningsをいかによく表現してまとめあげるか，そして簡単な指標をつくりあげるかの問題である。
この時は Reproducibility が大切なのである。

Scale Analyisis 等はこれに属する。

となる。

さて，数量化において実際的操作を行ふ立場は求めてゐることを統計数理的意味において信頼度高くしようとする立場である。

このため我々の数量化する行動に所謂 optimum をねらふ Rational Behaviour と言ふことが考へられねばならない。

これがある場合には相関比を最大にする，判断の成功率を最大にする， ある場合には安全性をねらふ Min-Max (Max-Mini) の下に思考するといふ形等々であらはれてきてゐるのである。

On the Confidence Interval of the Intersecting Point of two Regression Lines.

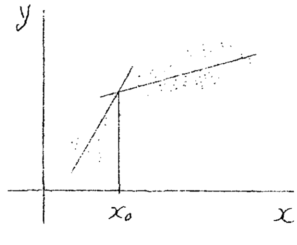
Chikio Hayashi

These problems often arise in practise, which are to require the turning points of phenomena.

By the idea of regression estimate in the theory of sampling, the confidence interval of the intersecting point is required.

For example, 2 regression lines are given.

$$\begin{cases} y_{1i} = \beta_1 x + \alpha_1 + e_{1i} \\ y_{2i} = \beta_2 x + \alpha_2 + e_{2i} \end{cases}$$



$$\begin{cases} E(e_{1i}) = E(e_{2i}) = 0 \\ E(e_{1i}^2) = E(e_{2i}^2) = \sigma^2 \\ E(e_{1i} e_{1j}) = E(e_{2i} e_{2j}) = 0 \qquad i \neq j \end{cases}$$

Then the estimate the intersecting point $x_0 = \frac{\alpha_2 - \alpha_1}{\beta_1 - \beta_2}$

is made from the data.

The mean square error of this estimate x' is τ^2.

$$\tau^2 = \sum_{i=1}^{2} \frac{(1-\rho_i)^2 \sigma y_i^2}{n_i} \left\{ \frac{1}{\sigma_{x_i}^2} + \left(1 + \frac{\overline{X}_i^2}{\sigma_{x_i}^2} + \frac{1}{n_i}\right) - 2 \frac{\overline{X}_i}{\sigma_{x_i}^2} \right\}$$

Where ρ_i is conelation coefficient.

σ_{x_i} is the variance of x

σ_{y_i} is the variance of y

$\bar{X}_i$ is the mean

n_i is the sample size

and the scales of x, y is given to in order that the relation

$$\alpha_2 - \alpha_1 = \beta_1 - \beta_2$$

holds.