Title	相関係数とmulticollinearity
Sub Title	Correlation coefficients and multicollinearity
Author	佐藤, 保
Publisher	慶應義塾経済学会
Publication year	1958
Jtitle	三田学会雑誌 (Keio journal of economics). Vol.51, No.7 (1958. 7) ,p.628(72)- 642(86)
JaLC DOI	10.14991/001.19580701-0072
Abstract	
Notes	資料
Genre	Journal Article
URL	https://koara.lib.keio.ac.jp/xoonips/modules/xoonips/detail.php?koara_id=AN00234610-19580701-0072

相関係数と multicollinearity

佐藤　保

一

経済分析ではしばしば重回帰による方法が用いられる。例えば所得階級別に資料をとり消費支出、可処分所得、家計人員の関係を求めるため、（消費支出＝C、可処分所得＝Y、家計人員＝m と書けば）

$$C=aY+bm+c \qquad (1)$$

として重回帰により、最小自乗法を使ってパラメーター、a、b、cを求める。そして(1)式の精度は重相関係数によって測られる。二変数間の関係を求める時、例えば

$$y=ax+b$$

の形ではこの方法で問題ないが、三変数以上になると複雑となる。そして

$$Z=a+bX+cY$$

において、「独立変数間に（X、Y間に）非常に密接な関係があるならば、充分な正確性をもって個々の回帰係数を求めることができなくなるであろう。しかしながら回帰方程式はもしXとYが与えられた場合のZを推測するための予測に対してはなお正当（或は有力）である。」[注1]と言われる。例として、「Zがコーヒーに対する需要であるとし、Xはコーヒーの価格、Yはコーヒーの生産国、例えばブラジルの国民所得とする。XとYとの間には確かに非常に密接な関係がある。何となれば国民所得の大部分はコーヒーの販売から得られるからである。このような状況の下では、XYとZの間の一次関係(linear relation)を決定することはできない。」[注2]このような状態を multicollinearity と言う。

ここで問題となる点を二つばかりあげてみよう。

(一)　XとYとの間に密接な関係があってもZを予測するためには使えるというが、この予測とは何であるか。

重相関係数が高ければ、回帰線からの誤差は少なくなるから、分析期間内、或は cross section では分析時期の各項目についてはよい予測が得られる。しかし分析期間を越えた次の期間のZを推定する、或いは次の時期の各項目の値を推定するという意味での予測では必ずしもよいとは言えない。最小自乗法の外に、パラメーター推定の方法として、構造推定と呼ばれる方法があるが、この方法によりパラメーターを求め重相関（この場合の重相関は最小自乗法による重相関とはやや意味を異にしよう。）として

$$R^2=\frac{\sum(z-\bar{z})^2-\sum(z-\hat{z})^2}{\sum(z-\bar{z})^2} \quad 或は \quad 1-\frac{\sigma_u^2}{\sigma_z^2}$$

（$\bar{z}$ は z の平均値、$\hat{z}$ は z の推定値、σ_u^2 は誤差分散、σ_z^2 は z の分散）

の形で求められる。構造推定によってパラメーターの値を求めた時は重相関は必ず低くなる。そこで分析期間内での予測という意味では最小自乗法に劣ることになるが、分析期間を越える期の予測についてはそう言えない。X、Yに密接な関係があるとき、正確なパラメーターの値を求めることができない。即ちX、Y、Zという経済量（変数）の間の正確な関係を求めることができないとすれば、そして何等かの条件を附して構造推定によって、より正確なX、Y、Zの関係、（或はa、b、cの値といってもよいであろう）が求められたとすれば、次期の予測においてはむしろ構造推定によるZの方が現実値に近いかもしれないのである。しかしこの場合次期として何期間をとるかということで問題が多い。

(二)　X、Y間に密接な関係があってはこまるというが、どの程度密接な関係がある場合パラメーターが不正確になるのか、その不正確という意味はどのようなことであるか。

まずX、Y間に完全に一次関係があれば、パラメーターa、b、cを求めることができなくなる。換言すれば変数としてX、Yどちらか一つでよいということである。この場合明らかに multicollinearity の状態にあると言ってよい。しかしこのような場合は、現実には存在しないであろう。それでは高い相関といってもどこで区切るかは分析者の判断による外ないということになる。そこでパラメーターの不正確さの判断が大きなウェイトをもつ場合が多いであろう。即ち経済変数間では、あらかじめ予想される関係（それは理論的根拠から乃至常識からと言ってもよい）があり、あらかじめ予想されるa、b、cの値の大きさ、符号に対して、それに反する値なり符号が計算された時は、そしてX、Yにかなり高い相関があれば、これより逆に multicollinear がありそのために符号が逆になったりしたのだと判断する。上の例で

$$Z=a+bX+cY$$

に於てbはマイナス、cはプラスの符号をもつと考えられるのでこれと違った符号が計算されればこまることになる。或は先の

$$C=aY+bm+c$$

での場合はa、bは共にプラスと考えられる。Cの代りに貯蓄S

をとって

$$S=a'Y+b'm+c'$$

とすれば a' はプラス、b' はマイナスと考えられる。しかしCとY、Cと m、SとY、Sと m は全部順相関の関係があるであろう。CとYが順相関、c と m が順相関であることから、a、b がプラスの符号をとることが一見して予想されるが、Sと m がそれだけを見れば順相関（それもかなり高い相関）があるにもかかわらず b' はマイナスになる。一見すると不思議に思えるが、重回帰係数の性質よりみれば、所得水準一定とすれば、家計人員と貯蓄は当然負の関係にあることが推察される。従って二変数間の相関がプラスであるにもかかわらずマイナスの符号が生じたからといって、multicollinear であるとはいえない。係数の符号については理論的根拠がある場合はそれに従って正確であるかどうかを判断しなければならない。理論的根拠が不明である場合は、x_1、x_2、x_3 の変数で

$$x_1=ax_2+bx_3+c$$

とし、x_1、x_2 が順相関であるにもかかわらず a がマイナスなら、x_3 を一定とした条件の下では x_1 と x_2 は逆相関の関係があるのだと判断する。そして標本数が多い場合は、x_3 がほぼ等しいグループにわけて何組か組をつくり、その中で x_1 と x_2 の関係を調べれば恐らく逆相関を示す組が多いであろう。そしてそれを合計すれば逆相関となるであろう。つまり x_1、x_2 の偏相関係数はマイナスであると判断されるのである。しかしこの場合でも常識と全くかけはなれたような値がでた時は x_1、x_2 間の関係について考慮することになるであろう。

ここで一つ考えておかなければならないのは今 x_1 を従属変数に x_2、x_3 を独立変数と考えた。経済分析の場合は特に x_2 と x_3 から x_1 を推定することが問題で、その限りでは独立変数 x_2、x_3 の相関のみが問題となるが、元来 multicollinearity の問題はR個の変数がある時その内 linear independent な変数がいくつあるかを問題とするものであり、その点で x_1、x_2、x_3 という変数があれば当然 x_2、x_3 の外に、x_1、x_2 及び x_1、x_3 の関係も問題となってくる。そうすれば

$$x_1=ax_2+bx_3+c$$

より独立変数 x_2、x_3 間の関係のみから a、b の値と符号を問題とすることとは異なってくるであろう。以下二三の具体例をあげてみよう。

二

三変数 x_1、x_2、x_3 があり、

$$x_1=ax_2+bx_3+c$$

なる当はめを考えよう。簡単化のため変数はすべて標準形 $\frac{x_i-\bar{x}}{\sigma_x}$（$\bar{x}$ は x の平均、σ_x は x の標準偏差）を考えておく。三変数の場合重相関係数の平方は次の式で与えられる。

$$R^2=\frac{r_{12}^2+r_{13}^2-2r_{12}r_{13}r_{23}}{1-r_{23}^2} \quad \text{(注3)}$$

又回帰係数は

$$a=-\frac{R_{12}}{R_{11}} \qquad b=-\frac{R_{13}}{R_{11}}$$

（ここで R_{11} は相関行列 r_{ij} における r_{11} の余因数を示す，R_{12}, R_{13} についても同様）

で与えられる。今 r_{12}、r_{13}（即ち従属変数と独立変数との間の関係）が与えられた時、独立変数間の関係 r_{23} が変化するにつけ、R^2 や回帰係数、その標準誤差等がどのように変化するかを考えてみよう。先ず r_{23} を未知数 x とし R^2 を y とかけば

$$y=\frac{r_{12}^2+r_{13}^2-2r_{12}r_{13}x}{1-x^2}$$

$$\frac{dy}{dx}=\frac{-2r_{12}r_{13}(1-x^2)-(r_{12}^2+r_{13}^2-2r_{12}r_{13}x)(-2x)}{(1-x^2)^2}=0$$

より

$$x=\frac{r_{12}^2+r_{13}^2\pm\sqrt{(r_{12}^2+r_{13}^2)^2-4(r_{12}r_{13})^2}}{2r_{12}r_{13}}$$

$$=\frac{r_{12}}{r_{13}} \text{ or } \frac{r_{13}}{r_{12}} \qquad (2)$$

で R^2 の極値を求めることができる。$R^2=1$ については

$$1=\frac{r_{12}^2+r_{13}^2-2r_{12}r_{13}x}{1-x^2}$$

より

$$x=\frac{2r_{12}r_{13}\pm\sqrt{(2r_{12}r_{13})^2-4(r_{12}^2+r_{13}^2-1)}}{2} \qquad (3)$$

として与えられる。$r_{12}=0.6$　$r_{13}=0.8$ の場合について考えて見よう。(2)式に代入すると

$$x=\frac{0.6}{0.8} \text{ or } \frac{0.8}{0.6}$$

$$x=\frac{4}{3} \text{ or } \frac{3}{4}$$

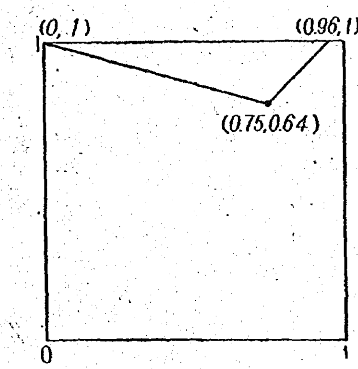
(3)式より

$$x=\frac{2\cdot0.48\pm\sqrt{(0.96)^2-0}}{2}$$

$$x=0.96 \text{ or } 0$$

を得る。$x=0.75$ のときの R^2 を求めると 0.64 となる $R=0.8$。これは x_1 と x_3 の単純相関係数に等しい。即ち x_2 を附け加えた効果はないといってよい。大体は第一図の如くなるであろう。即ち $r_{23}=0.75$ 以下では独立変数間の相関が低ければ低い程重相関は高くなり、従って分析期間内の予測は確からしくなるが、0.75 を越えると、独立変数間の相関が高くなる程重

第一図



相関係数は高くなる。その場合の回帰係数を見ると、先ず相関行列をつくると左の如くなる。$r_{23}=0.75$ として a、b を求めて見よう。

$$\begin{vmatrix} r_{11} & r_{12} & r_{13} \\ r_{21} & r_{22} & r_{23} \\ r_{31} & r_{32} & r_{33} \end{vmatrix} \quad \begin{vmatrix} 1 & 0.6 & 0.8 \\ 0.6 & 1 & r_{23} \\ 0.8 & r_{23} & 1 \end{vmatrix}$$

$$R_{11}=\begin{vmatrix} 1 & 0.75 \\ 0.75 & 1 \end{vmatrix}=1-(0.75)^2=0.4375$$

$$-R_{12}=\begin{vmatrix} 0.6 & 0.75 \\ 0.8 & 1 \end{vmatrix}=0 \quad \text{即ち} \quad r_{12}=r_{13}r_{23}$$

$$-R_{13}=\begin{vmatrix} 1 & 0.6 \\ 0.75 & 0.8 \end{vmatrix}=0.8-0.45=0.35$$

$$a=0 \qquad b=0.8$$

$a=0$ となるから結局

$$x_1=bx_3x_1=0.8x_3$$

となる。そして 0.75 を境にして、それ以上 r_{23} が高い値をとるときは a はマイナスとなることがわかる。そこで先の例で従属変数を消費支出と貯蓄と考え、x_2 を家計人員、x_3 を可処分所得と考えれば、従属変数が食費支出であるときは r_{23} が 0.75 以下、従属変数が貯蓄であるときは 0.75 以上であることが要求されるであろう。[注4] 回帰係数の標準誤差についても、分散を使ってガウスの乗数を計算すると、標本数 n とすると

$$C_{11}=C_{22}=\frac{1}{n}\cdot\frac{1}{1-r_{23}^2}$$

誤差分散　$S^2=1-R^2$

これより回帰係数の標準誤差は

$$S_a=S_b=\frac{\sqrt{1-R^2}}{\sqrt{(1-r_{23}^2)}\cdot\sqrt{n}}$$

一方回帰係数は、例えば a は

$$a=\frac{r_{12}-r_{23}r_{13}}{1-r_{23}^2}$$

で与えられるから、標準誤差との比率をとれば

$$\frac{a}{\sqrt{C_{11}}\cdot\sqrt{S^2}}=\frac{r_{12}-r_{23}r_{13}}{\sqrt{1-r_{23}^2}}\cdot\frac{\sqrt{n}}{\sqrt{1-R^2}}$$

となって、第一に標本数が問題であるが、これを一定とすれば、r_{23} が高い場合小さいとはいえない。この例で $r_{23}=0.95$ と $r_{23}=0.12$ とを比較してみると、

$r_{23}=0.95$ の場合

$R^2=0.9026 \quad S^2=(1-R^2)=0.0974 \quad S=0.3121$

$$C_{11}=\frac{1}{1-r_{23}^2}\cdot\frac{1}{n}=\frac{1}{0.0975}\cdot\frac{1}{n}=\frac{10.256}{n} \qquad \sqrt{C_{11}}=\frac{3.202}{\sqrt{n}}$$

$a=-1.641 \qquad S\sqrt{C_{11}}=0.9993$

$$\left|\frac{a\sqrt{n}}{S\sqrt{C_{11}}}\right|=1.64$$

$r_{23}=0.12$ の場合

$R^2=0.8977$

$S^2=1-0.8977=0.1023 \quad S=0.3198$

$$C_{11}=\frac{1.0146}{n} \qquad \sqrt{C_{11}}=\frac{1.007}{\sqrt{n}} \qquad \frac{\sqrt{C_{11}}\cdot S}{\sqrt{n}}=\frac{0.3320}{\sqrt{n}}$$

$a=0.511$

$$\frac{a\sqrt{n}}{\sqrt{C_{11}}\cdot S}=1.51\sqrt{n}$$

となり、独立変数間の相関が高い時でも回帰係数の標準誤差が相対的に大きくなるとはいえない。しかし係数 a のマイナスの値が、理論的に1を越えないと考えられるならば当然独立変数間の相関が高いために係数が不正確になったと考えられる。この点は微妙な所である。今の場合、x_2、x_3 を独立変数として固定し、従属変数 x_1 を推定する時のことを考えたが、即ち x_2、x_3 に誤差がなく x_1 方向のみの誤差を考え、回帰係数の符号を問題にした。x_1、x_2 を原因とし x_3 が結果であるときはそれでよいが、単に三者の関係という時はこれだけでなくあらゆる方向の誤差が問題にされなければならないであろう。次にそのことについて考察しよう。

三

x_1、x_2 の二変数があり、x_1 を従属変数、x_2 を独立変数とすれば、通常の最小自乗法では x_2 には誤差なく、x_1 のみに誤差があると考える。そして

$$x_1=ax_2+b$$

で a、b はそれぞれ x_1 方向に誤差をとることで計算される。x_2 が原因で x_1 が結果の場合はこれでよいが、単に両方の関係というときにはこの逆も考えられる。即ち

$$x_2=a'x_1+b'$$

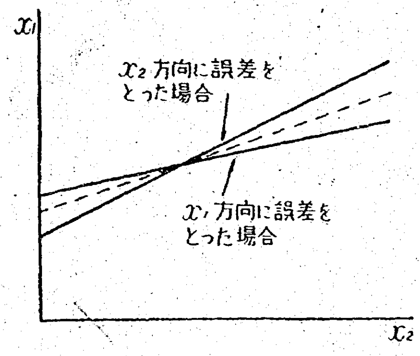
で a'、b' はそれぞれ x_2 方向に誤差をとることによって計算される。

$$a=\frac{\sum(x_1-\bar{x}_1)(x_2-\bar{x}_2)}{\sum(x_2-\bar{x}_2)^2}$$

$$a'=\frac{\sum(x_1-\bar{x}_1)(x_2-\bar{x}_2)}{\sum(x_1-\bar{x}_1)^2}$$

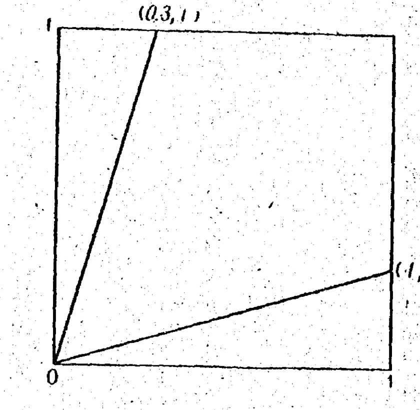
であるが、変数を標準形にして考えれば $a=a'=r_{12}$ となる。第二図の如く通常 x_1 方向に誤差をとったときと、x_2 方向に誤差をとったときと、二本の回帰線がひける。両者の関係という点ではこの二本の線共に満足すべきものでなく、その間の図の点線で示された線の如きものが、x_1、x_2 間の真の関係に近いのではないかと考えられよう。二本の回

第　二　図



帰線が一致するのは相関係数が1、即ち完全な直線関係にあるときのみである。例えば相関係数0.3として、標準形について考えると第三図の如くなろう。

第三図



x_1方向に誤差をとれば

$$x_1=0.3x_2$$

x_2方向に誤差をとれば

$$x_2=0.3x_1 \qquad x_1=\frac{1}{0.3}x_2$$

となる。この図を bunch と言う。そして各線を beam と言う。相関係数0.3では beams の間が相当ひらいているが、相関が高まるにつれて接近し、相関係数1のとき一本になる。従ってこの場合は相関係数が高ければ高い程、即ち beams の間のひらきが小さければ小さい程、x_2からx_1の変動を説明することが確か（同様にx_1からx_2を説明することが確か）となる、即ち回帰係数の信頼性が増すことになろう。このような考え方を三変数以上にも推し進めることができる。一般にx_1……x_kのk個の変数がありx_1方向に誤差をとるとき、標準形を考えれば

$$R_{11}x_1+R_{12}x_2+R_{13}x_3+\cdots\cdots+R_{1k}x_k=0$$

なる関係があり

$$x_1=a_2x_2+\cdots\cdots+a_kx_k$$

とすれば

$$a_i=\frac{-R_{1i}}{R_{11}}$$

で表わされる。

三変数の場合、各方向の誤差を考えれば、

$$R_{11}x_1+R_{12}x_2+R_{13}x_3=0 \qquad x_1\text{ 方向の誤差}$$

$$R_{21}x_1+R_{22}x_2+R_{23}x_3=0 \qquad x_2\text{ 方向の誤差}$$

$$R_{31}x_1+R_{32}x_2+R_{33}x_3=0 \qquad x_3\text{ 方向の誤差}$$

これより各方向に誤差をとった時の回帰係数を比較するため、

$$x_1=\frac{-R_{12}}{R_{11}}x_2+\frac{-R_{13}}{R_{11}}x_3$$

$$x_1=\frac{-R_{22}}{R_{21}}x_2+\frac{-R_{23}}{R_{21}}x_3$$

$$x_1=\frac{-R_{32}}{R_{31}}x_2+\frac{-R_{33}}{R_{31}}x_3$$

の形に直せば、x_1とx_2、x_1とx_3について、各方向に誤差をとったときの回帰係数を比較することができる。そして分子の値を縦軸に、分母の値を横軸にとって一点を定め、原点と結べば一本の beam ができる。x_1とx_2について三本の beams ができるから、これより一つの bunch を書くことができる。x_1とx_3についても同様に、又x_2

を左辺の従属変数としてとることによってx_2とx_3の間の回帰係数＝bunch を書くことができる。結局三変数の場合は、二変数間の bunch 三個と、三つめの変数を加えたときの bunch 三個と合計六個の bunches ができるがこれを総称して bunch map ということになる。より一般的に言えば標準化した変数で

$$a_1x_1+a_2x_2+\cdots\cdots+a_kx_k=0$$

においてi、jをこれら諸変数の添字の中の二つであり$i<j$であるとする。

$$x_i=a_{ij}x_j+\cdots\cdots$$

としこの二変数間についてx_γ方向に最小自乗法を使ったときの回帰係数は

$$a_{ij}^{(\gamma)}=-\frac{R_{\gamma j}}{R_{\gamma i}}$$

$$\gamma=1,2,\cdots\cdots k$$

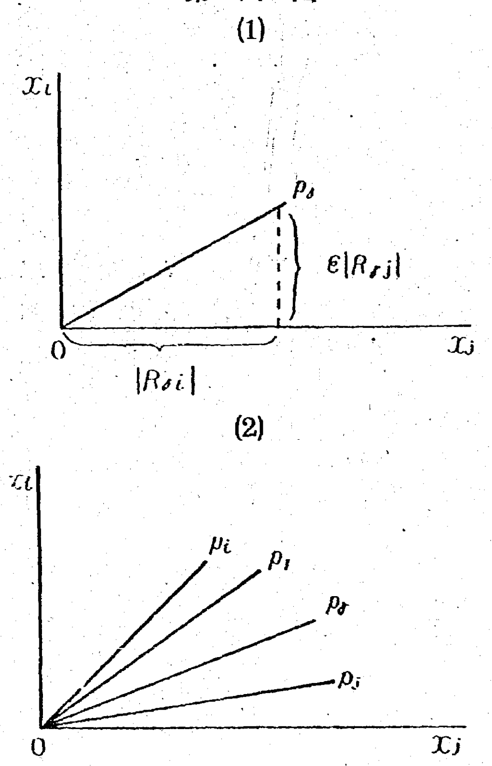
で示される。そしてx_iを縦軸に、x_jを横軸、($i<j$)にとったx_i、x_j平面上で座標$|R_{\gamma i}|$, $\varepsilon|R_{\gamma j}|$をもつ点$p_\gamma$をしるし、原点0から点$p_\gamma$に向う beam o.p.r を描く、ここで$\varepsilon$は

$$\varepsilon=\begin{cases}+1 & R_{\gamma i}\text{ と }R_{\gamma j}\text{ が異符号であるとき}\\ -1 & R_{\gamma i}\text{ と }R_{\gamma j}\text{ が同符号であるとき}\end{cases}$$

と定義される（第四図参照）。

k個の変数があれば一つの bunch にk個の beams が描かれる。

第四図



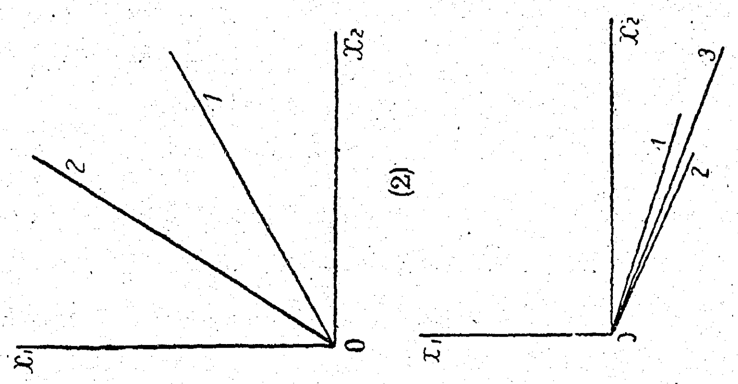
そして最初（x_1、x_2）bunch をしらべ次に（x_1、x_3）bunch を調べるというように進んでゆき、各 bunch 例えば（x_i、x_j）bunch は、その bunch がそこで構成されている変数から一変数を除いた（x_i、x_j）bunch と比較されることになる。即ち単に（x_1、x_2）よりなる bunch と、x_3を加えたときの（x_1、x_2）bunch とを比較して beams の状態を見る。そこから基本的に次の三つの基準で新変数導入の効果を測る。

(1) Useful 新変数を導入したことが良いと判断する。

誤差の方向をどの変数にとっても回帰係数の値が変らなければ、各 beam は一つの方向に揃うことになる。

例えば今変数x_1の変数x_2への回帰係数を問題にしている時、x_1、x_2のみを考えているときは(1)図の如くなり（beam についている番号

第五図



は誤差の方向を示す)、beams が相当ひらいているとしよう。そして新変数 x_3 を導入したときは、(2)図の如く beams が一致してきたとすれば、新変数の導入は有効であったと考えられるであろう。即ち新変数を加えることは必要であったのである。useful の基準として三つがあげられている。

(a)(注5) 新しい bunch の beams がもとの bunch の beams に比べて著るしく引き締まってくる。

(b) 新変数の beams が他の beams の存在範囲の内側に入ってくる。

(c) 新しい bunch における beams の全体的方向がもとの bunch における beams に比べて、大きく変化する。

(2) Superflous 新変数の導入は有害ではないが必要でもない。

(a) bunch の beams が新変数を導入しても目立ってそろってこない。

(b) bunch の beams の全体的方向がほとんど変化しない。

(c) 新変数の beam が、他の beams の定める存在範囲の外側に落ちる。

(d) 新変数の beam が、新しい bunch において他の beams よりも著しく短い。

(e) 他の変数の beams が、新変数の導入によって著しく短くはならない。

ここで bunch map analysis では beams のひらき方と共に beams の長さが問題となるのである。x_1……x_k の変数があり、x_1 方向に誤差を考えるとき、x_1 を除いた他の諸変数間によりよい線型関係があればある程 R_{11} は小さい値をとる。又 R_{12}、R_{13}、……R_{1k} も大体それに順じて小さい値をとるので、x_1 を除いた残りの変数がよりよい線型関係にあればある程 x_1 方向の beam は各 bunch において小さいことになる。即ち beam の長さが短いということは、その変数を入れても、あまり線型関係を改善しないということを意味している。

(3) Detrimental 新変数を導入することは有害である。

新変数を導入したにもかかわらず bunch の beams がもとの bunch の beams よりもより大きい拡がりを示すならばその新変数を有害と判断する。従って multicollinear と考える。

以上のことはいずれも、うなずけることであるが、実際の分析となると、どの程度 beams が揃えば useful と判断し、どの程度をあまり変らないとして superflous とし、どの程度ひらけば detrimental とするかは、分析者の判断によるのであって、客観的規準はない。ここに buncp map analysis の欠点があるといえばいえるが、又融通性もあるといえる。又三変数では6個の bunches ができることを述べたが、変数が増えるに従って bunch の数は急速に増してゆく、そのいずれの bunch においても、とりあげられた変数(x_1、x_2)に対して残りの変数の効果が useful であるとは限らないであろう。そこでつくられた bunch についてそれぞれ判断して useful と思われるものが多ければ、全体として良いと判断する外ないであろう。その場合でも全体の何割以上が useful であれば良いという規準はない。このことを一目でわかるために star map と呼ばれる表をつくる。これは上の如く表をつくり、変数の組合せをつくる。そして例えば変数1、2(x_1、x_2)に3を加えたとき、useful なら*を、1、3に2を加えたとき superflous なら●を、2、3に1を加えたとき detrimental なら○を、というように、各項目に判断の印を記入してゆき、結局*印が多ければ全体として、この変数の組は、真の回帰係数を求めるのに良い条件をそなえていると判断するのである。以上が bunch map analysis のごく概略であるが、主観的判断を含む故に困難を伴う場合も起ってくる。二、三の例を示してみよう。

追加される変数	回帰係数 1, 2	1, 3	2, 3	1, 4	…
1			○		
2		●			
3	*				
1 2 4					
……					

先にあげた $r_{12}=0.6$, $r_{13}=0.8$, $r_{23}=0.95$ の場合と、$r_{12}=0.6$, $r_{13}=0.8$, $r_{23}=0.12$ の場合について考えよう。この場合重相関係数はほぼ等しいのであった。手続きの順序として、相関行列をつくる。$r_{12}=0.6$, $r_{13}=0.8$, $r_{23}=0.95$ については左の如くなる。これから二変数ずつの組合せを考えると、

$$\begin{vmatrix} 1 & 0.6 & 0.8 \\ 0.6 & 1 & 0.95 \\ 0.8 & 0.95 & 1 \end{vmatrix} \quad \begin{vmatrix} r_{11} & r_{12} \\ r_{21} & r_{22} \end{vmatrix} \quad \begin{vmatrix} r_{11} & r_{13} \\ r_{31} & r_{33} \end{vmatrix} \quad \begin{vmatrix} r_{22} & r_{23} \\ r_{32} & r_{33} \end{vmatrix}$$

次ぎに余因数の行列式をつくる。

$$\begin{vmatrix} R_{11} & R_{12} \\ R_{21} & R_{22} \end{vmatrix} \quad \begin{vmatrix} R_{11} & R_{13} \\ R_{31} & R_{33} \end{vmatrix} \quad \begin{vmatrix} R_{22} & R_{23} \\ R_{32} & R_{33} \end{vmatrix}$$

三変数については

$$\begin{vmatrix} R_{11} & R_{12} & R_{13} \\ R_{21} & R_{22} & R_{23} \\ R_{31} & R_{32} & R_{33} \end{vmatrix}$$

この例では

$$\begin{vmatrix} 1 & -0.6 \\ -0.6 & 1 \end{vmatrix}$$

$$\begin{vmatrix} 1 & -0.8 \\ -0.8 & 1 \end{vmatrix}$$

$$\begin{vmatrix} 1 & -0.95 \\ -0.95 & 1 \end{vmatrix}$$

$$\begin{vmatrix} 0.0975 & 0.16 & -0.23 \\ 0.16 & 0.36 & -0.47 \\ -0.23 & -0.47 & 0.64 \end{vmatrix}$$

となる。又 $r_{12}=0.6$, $r_{13}=0.8$, $r_{23}=0.12$ では

$$\begin{vmatrix} 1 & 0.6 & 0.8 \\ 0.6 & 1 & 0.12 \\ 0.8 & 0.12 & 1 \end{vmatrix}$$

となり、これより余因数の行列式をつくれば、

$$\begin{vmatrix} 1 & -0.6 \\ -0.6 & 1 \end{vmatrix}$$

$$\begin{vmatrix} 1 & -0.8 \\ -0.8 & 1 \end{vmatrix}$$

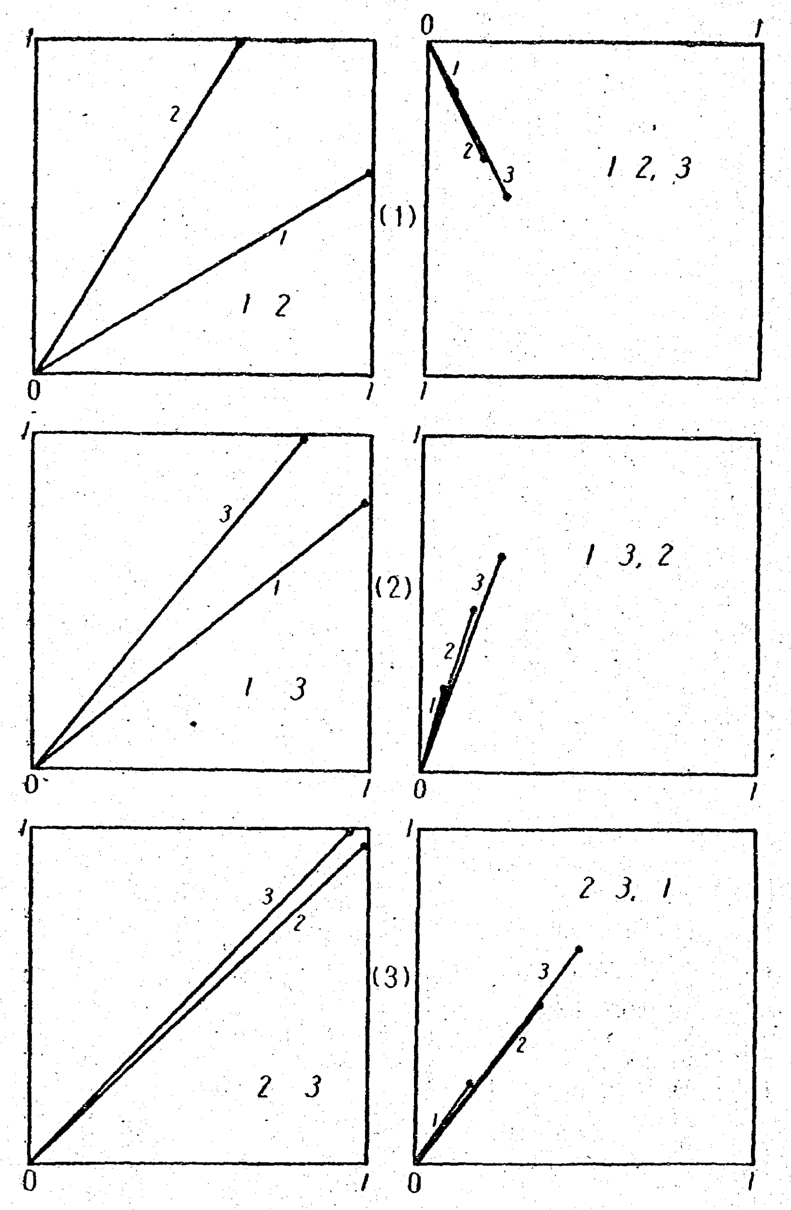
$$\begin{vmatrix} 1 & -0.12 \\ -0.12 & 1 \end{vmatrix}$$

$$\begin{vmatrix} 0.9856 & -0.504 & -0.728 \\ -0.504 & 0.36 & 0.36 \\ -0.728 & 0.36 & 0.64 \end{vmatrix}$$

となる。前の規則に従って bunch map をつくると次のようになる(第六・七図参照)。

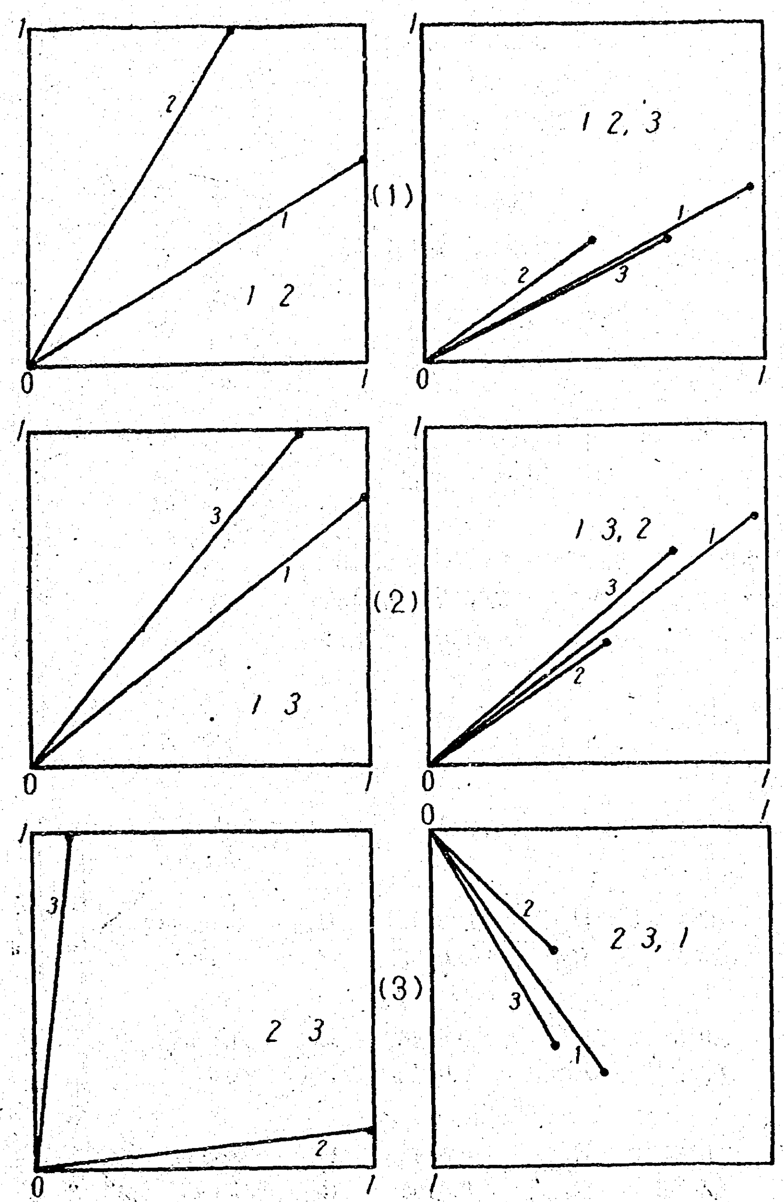
第六図をみると各 bunch 共 beams が非常によくそろってきており、この限りでは(1)(2)(3)共に useful と判断されそうである。ただ(3)については新たに加えられた1の beam が短いという点で superflous と判断されるかもしれない。この点は微妙なところである。しかしたとえ superflous と判断されてもこれ一つだけから全体を排除することはできないであろう。第七図についてもすべて useful と判断されよう。(2)について2 beam がやや短いが、たいして問題とはならないであろう。star map をつくれば、次の如くなろう。その判定は大きな差がない。

第六図



$(r_{12}=0.6 \quad r_{13}=0.8 \quad r_{23}=0.95)$

第七図



$(r_{12}=0.6 \quad r_{13}=0.8 \quad r_{23}=0.12)$

(1)

	1, 2	1, 3	2, 3
1			* or ●
2		*	
3	*		

(2)

	1, 2	1, 3	2, 3
1			*
2		*	
3	*		

したがって2、3を独立変数と考えたとき、高度の相関があるときとないときでも、star map に記された結果はあまり変りのないことになる。それではどのようなとき detrimental になるであろうか。次の例を考えよう。

$r_{12}=0.34, r_{13}=0.31, r_{23}=0.95$ の場合と、$r_{12}=0.34, r_{13}=0.31, r_{23}=0.73$ の場合についてみる。変数2、3を独立変数として、先ず極値を考えると $0.31/0.34=0.911$ でこのときの R^2 は $0.34^2=0.1156$ となることがわかろう。又 $R^2=1$ なる点を求めるため前の公式に代入すれば、

$$=\frac{2\cdot 0.34\cdot 0.31\pm\sqrt{(2\cdot 0.34\cdot 0.31)^2-4(0.34^2+0.31^2-1)}}{2}$$

$$=0.995 \text{ or } -0.7887$$

$r_{23}=0.95$ のときのR^2は0.117であり、$r_{23}=0.73$ のときのR^2は0.124でわずかに高い。bunch map をつくってみよう。相関行列と、余因数の行列を示すと次のようになる。

$$\begin{vmatrix} 1.00 & 0.34 & 0.31 \\ 0.34 & 1.00 & 0.73 \\ 0.31 & 0.73 & 1.00 \end{vmatrix} \qquad \begin{vmatrix} 1.00 & 0.34 & 0.31 \\ 0.34 & 1.00 & 0.95 \\ 0.31 & 0.95 & 1.00 \end{vmatrix}$$

$$\begin{vmatrix} 1.00 & -0.34 \\ -0.34 & 1.00 \end{vmatrix} \qquad \begin{vmatrix} 1.00 & -0.34 \\ -0.34 & 1.00 \end{vmatrix}$$

$$\begin{vmatrix} 1.00 & -0.31 \\ -0.31 & 1.00 \end{vmatrix} \qquad \begin{vmatrix} 1.00 & -0.31 \\ -0.31 & 1.00 \end{vmatrix}$$

$$\begin{vmatrix} 1.00 & -0.73 \\ -0.73 & 1.00 \end{vmatrix} \qquad \begin{vmatrix} 1.00 & -0.95 \\ -0.95 & 1.00 \end{vmatrix}$$

$$\begin{vmatrix} 0.4671 & 0.1137 & -0.0618 \\ 0.1137 & 0.9039 & -0.6246 \\ -0.0618 & -0.6246 & -0.8844 \end{vmatrix} \qquad \begin{vmatrix} 0.0975 & -0.0455 & 0.013 \\ -0.0455 & 0.9039 & -0.8446 \\ 0.013 & -0.8446 & 0.8844 \end{vmatrix}$$

三変数のときの bunch の beam 計算の為の数値を書けば bunch map は次の如くなる。

$$\frac{R_{12}}{R_{11}}=\frac{-0.0455}{0.0975} \qquad \frac{R_{22}}{R_{21}}=\frac{0.9039}{-0.0455} \qquad \frac{R_{32}}{R_{31}}=\frac{-0.8446}{0.013}$$

$$\frac{R_{13}}{R_{11}}=\frac{0.013}{-0.0975} \qquad \frac{R_{23}}{R_{21}}=\frac{-0.8446}{-0.0455} \qquad \frac{R_{33}}{R_{31}}=\frac{0.8844}{0.013}$$

$$\frac{R_{13}}{R_{12}}=\frac{0.013}{-0.0455} \qquad \frac{R_{23}}{R_{21}}=\frac{-0.8446}{-0.0455} \qquad \frac{R_{33}}{R_{31}}=\frac{0.8844}{0.013}$$

$$\frac{R_{12}}{R_{11}}=\frac{-1.1137}{0.4671} \qquad \frac{R_{22}}{R_{21}}=\frac{0.9039}{-0.1137} \qquad \frac{R_{32}}{R_{31}}=\frac{-0.6246}{-0.0618}$$

$$\frac{R_{13}}{R_{11}}=\frac{-0.0618}{0.4671} \qquad \frac{R_{23}}{R_{21}}=\frac{-0.6246}{-0.1137} \qquad \frac{R_{33}}{R_{31}}=\frac{0.8844}{-0.0618}$$

$$\frac{R_{13}}{R_{12}}=\frac{-0.0618}{-0.1137} \qquad \frac{R_{23}}{R_{22}}=\frac{-0.6246}{0.9039} \qquad \frac{R_{33}}{R_{32}}=\frac{0.8844}{-0.6246}$$

第八図について見れば(1)は superflous (2)は useful 或いは superflous (3)は superflous 或いは detrimental と判断されよう。これに対して第九図では(1)(2)(3)共に detrimental と判断されよう。判断の基準を甘くすれば(2)(3)は superflous と判断されるかもしれない。図から判断すれば、beam がひらいていないということから第八図の方が良いと判断されよう。bunch map では beams がひらいてしまう、即ちこの場合では三本の beams の方向（プラスかマイナス）が一致しないといけないわけである。今 r_{12}、r_{13}、r_{23} 共にプラスの値と考えると、三変数の余因数の行列式で、R_{11}、R_{22}、R_{23} は必ずプラスの値をとるが、

$$R_{12}=-\begin{vmatrix} r_{12} & r_{23} \\ r_{13} & 1 \end{vmatrix}=-(r_{12}-r_{13}r_{23})$$

$$R_{13}=\begin{vmatrix} r_{12} & 1 \\ r_{13} & r_{23} \end{vmatrix}=r_{12}r_{23}-r_{13}$$

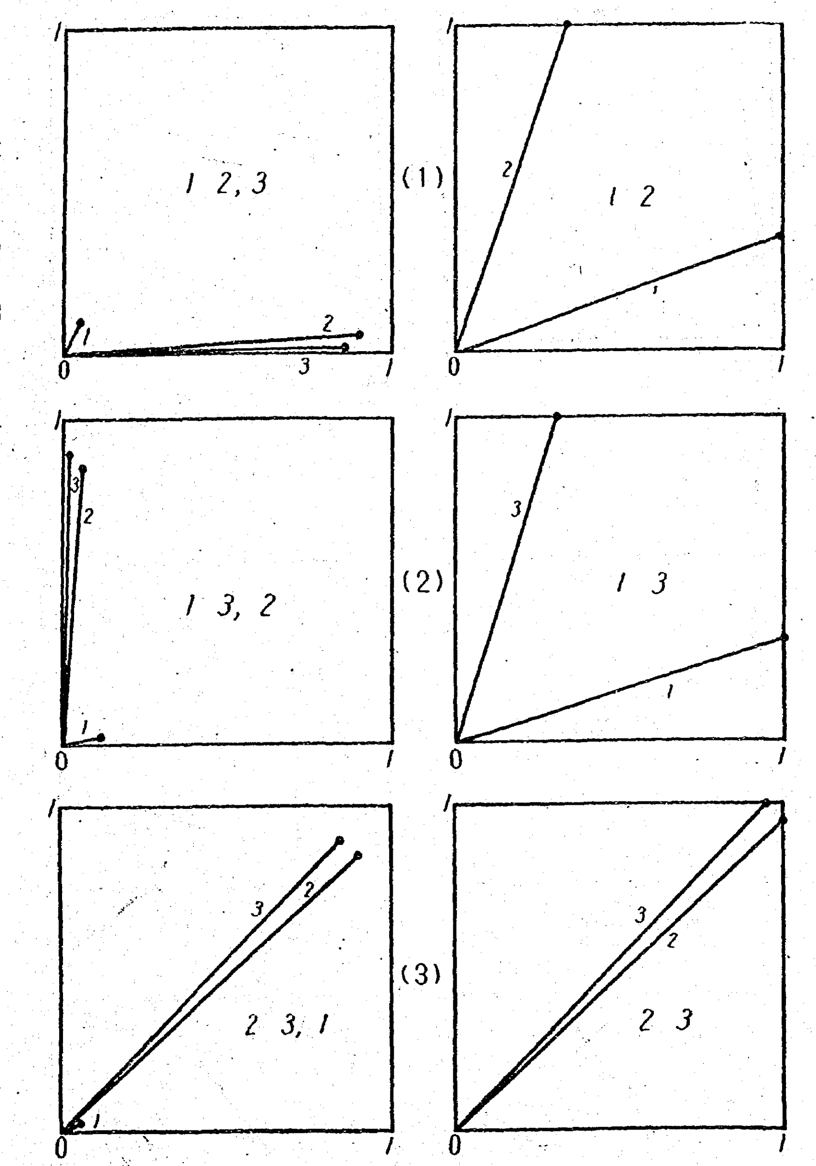
$$R_{23}=-\begin{vmatrix} 1 & r_{12} \\ r_{13} & r_{23} \end{vmatrix}=-(r_{23}-r_{12}r_{13})$$

でこれらの値は r_{12}、r_{13}、r_{23} の種々の値に従ってプラスになったり、マイナスになったりする、三変数の場合六通り考えられる

$R_{12}+$ $R_{13}-$ $R_{23}+$	$R_{12}-$ $R_{13}+$ $R_{23}-$
$R_{12}+$ $R_{13}+$ $R_{23}-$	$R_{12}-$ $R_{13}-$ $R_{23}+$
$R_{12}-$ $R_{13}-$ $R_{23}-$	$R_{12}+$ $R_{13}+$ $R_{23}+$

が、右側の三通りの場合、beams がひらくことがわかる。以上の概要により回帰係数の有効性を確かめる上に bunch map の方法が有効であることがわかるが、主観的判断の要素が多いことに不安定性はまぬがれない。この他 multicollinearity の判断に weighted regression の方

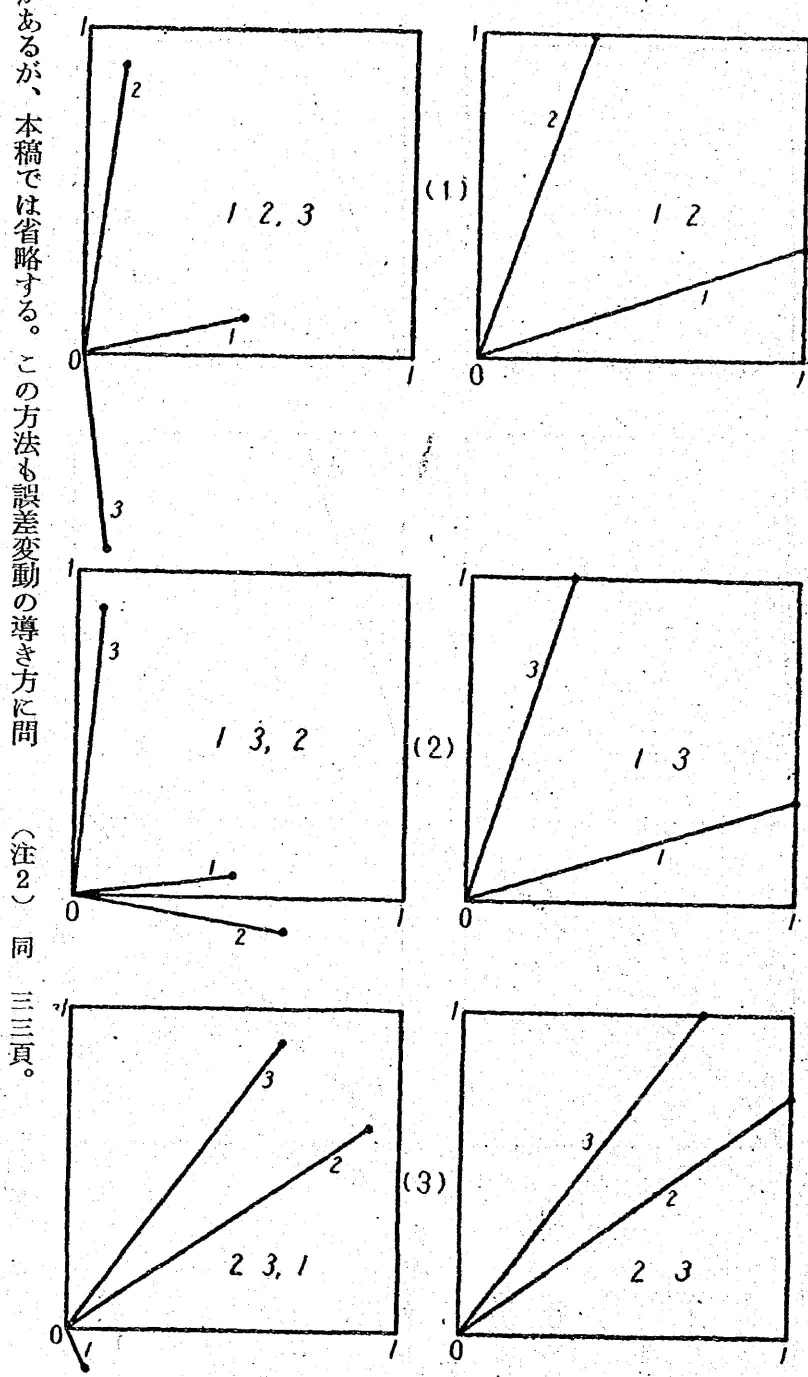
第八図



相関係数と multicollinearity

$(r_{12}=0.34 \quad r_{13}=0.31 \quad r_{23}=0.95)$

第九図



($r_{12}=0.34$ $r_{13}=0.31$ $r_{23}=0.73$)

法があるが、本稿では省略する。この方法も誤差変動の導き方に問題があり、回帰分析では、回帰係数の有為性(回帰係数0という仮説がすてられること)が第一の重要な問題であって、たとえ multi-collinearity でないと判断されても、回帰係数が有為でなければ、その式が良いといえないことは当然のことであろう。

(注1) G. Tintner Econometrics 三三頁。

(注2) 同　三三頁。

(注3) スネデカー「統計的方法　下」畑村又好、津村善郎、奥野忠一、田中祐輔訳、三四〇頁。

(注4) 同　三五八頁。

(注5) 日本経済分析研究会、消費部会研究報告書、一一三頁―一一七頁。

パンカースト著

『サン・シモン主義者ミルおよびカーライル――近代思想序説』

The Saint Simonians Mill and Carlyle
A Preface to Modern Thought
by Richard Pankhurst

"in short, although I am not St. Simonist, nor at all likely to become one, *je tiens bureau de St. Simonisme chez moi*"—John Stuart Mill—

一

イギリスの社会思想史や経済学説史を読んで、いつも考えさせられることは、この国の偉大な思想家や経済学者の多くが、たとえば大学教授のような学問の研究を職業とする人々ではなかったということである。アダム・スミスの如きはわずかにその例外にすぎない。リカードウやマルサス、ジェレミー・ベンサムとジェームズ・ミルそしてその息子ジョン・スチュアート・ミルさらにトーマス・カーライル、ロバート・オーエンと、英国思想史上の人々を想いうかべるならば、この国における思想家や理論家たちが、ただ象牙の塔にたてこもり、静かな書斎の一室にとじこもってその理論をうち樹てたものではなかったことに気がつくであろう。なるほど、ジェレミー・ベンサムの如きは孤独を愛し、名声嘖々たるなかにありながら、ジェームズ・ミルやリカードウ等のごく限られた範囲の人々としか交際しなかったいわゆる孤高の哲学者であった。しかしそれにもかかわらず、その「最大多数の最大幸福」の理想は勃興する英国資本主義のイデオロギーとして実践的な役割を果し、十九世紀前半から後半にかけての英国思想界を風靡し、ダイシーの言葉をかりるならば、「ベンサム主義もしくは個人主義の時期」[1](The Period of Benthamism or Individualism)を現出せしめた。

だがわれわれにとってもっとも興味深いのは、ベンサムの後継者として出発し、正統派経済学の巨匠たるべく運命づけられながら、晩年になるにつれて次第に功利主義に懐疑的となり、自由放任主義に批判的になるに至った過渡期の思想家、「さまよえる経済学者」ジョン・スチュアート・ミルの英国社会思想史上における地位であろう。一体、ミルはたんなる自由主義者として終始したのか、それとも社会主義に改宗してこの世を去ったのか、この問題は、必ずしも明らかではない。

しかしいずれにせよ、彼が社会主義に異常な関心を抱いていたことは、その遺稿「社会主義論」[2]をみても明らかであろう。とくにフランスのサン・シモンの社会主義からつよい刺戟をあたえられ、それから非常に多くのものを学んだことが、その自叙伝のなかに記さ