

## 多変量解析 事例集

曾谷勝義

NEC ソフトウェア関西

### 1. はじめに

多変量解析(Multivariate Analysis)は統計学理論の全領域を網羅していると考えても良い。個体の1つ又は数個の標本に関して行われる多変数測定分析としての統計学分野である。「多変量の分析はn個の実験材料に関するp個の測度の結合性に関するものである。」(Kendall)

理論上は、多変量の解析はp個の平均、及びp個の分散に払うのと同じような注意を、変量間の $p(p-1)/2$ 個の異なる共分散を研究する事にある。「多変量な統計学が基礎にしている理論は、数学モデルとして多変量正規分布(m.n.d.)を仮定している。この上で3つの統計量(平均、分散、共分散)は、それぞれパラメータとなっている。このパラメータの推定、検定を行うことにある。」(Anderson)

これらのことは、多変数測度に対する多変量正規分布の仮定が、1変量測定に対する正規性の仮定を導くのと同じ中心極限定理に起因する。この多変量正規分布が、数学的に行列法で取り扱い易い点が、計算理論を大きく進展させた。その計算結果の推定法、検定法が有力である点が魅力的となっている。

この多変量解析は、互いに従属な群の変量の関係を研究するものであるが、これには従属関係(dependence)か、又は相互依存関係(interdependence)のいずれかを扱うことになる。従属関係とは、問題の条件により1つ又は2つ以上の変量に対して、他の変量が従属している関係を調べる事にある。これは回帰とか要因実験等が上げられる。相互依存関係とは、群の変量とそれらの相互関係を調べる事にある。相関を始め、成分や因子の分析等が上げられる。

多変量解析の理論関係は、今日いろいろな書籍や研究文献に多く出ている。今回の多変量解析事例集は、与えられた事象に関して、如何にして分類し、多変量としての考え方を進めて行くか、また多変量解析をいかにうまく当てはめて行くかを主に述べて、数式は必要最小限とし、必要が有る場合のみ少し記載したい。

## 2. 解析の前提

解析対象として2群の変数を考える。第一群を $\{x_1, \dots, x_p\}$ 、第二群を $\{y_1, \dots, y_q\}$ とおく。一般的に  $p$  個の説明変数(predictor variables)に対して、 $q$  個の目的変数(criterion variables)に関する情報を与えようとする試みは、統計学の面で多くの研究が行われている。この目的変数及び説明変数は、いくつかの尺度(scale)の概念で測られる。分類尺度(nominal scale)、順序尺度(ordinal scale)、間隔尺度(interval scale)、比尺度(ratio scale)等がある。いま代表的な重回帰分析を取り上げる。

目的変数を  $y$  で表し、説明変数を  $x_1, \dots, x_p$  で表すと、 $y$  の変動の中で  $x_1, \dots, x_p$  で説明出来ない変動を  $\varepsilon$  とおく。この  $\varepsilon$  は一般に残差(residual)と呼ばれる。説明変数から構成される関数を  $f(x_1, \dots, x_p)$  とおく。この関数  $f(x_1, \dots, x_p)$  が滑らかであると、それぞれある基準値  $x_1 = m_1, \dots, x_p = m_p$  の回りに Taylor 展開して近似しているものとする事が出来る。この  $\varepsilon$  は Taylor 展開2次以上のものをすべて含めた誤差(error)と考える事も出来る。この時の残差  $\varepsilon$  は、 $n$  個のサンプル、 $p$  個の説明変数に関して、各測定値間の残差として  $\varepsilon_i$  ( $i = 1, 2, \dots, n$ ) とおく。

重回帰分析においては、この残差  $\varepsilon_i$  ( $i = 1, 2, \dots, n$ ) に関して、以下の4つを基礎的条件として仮定する。

(1)独立性の仮定：

$$\varepsilon_i \text{ と } \varepsilon_{i'} \text{ とは互いに独立である。} \quad (i \neq i', \quad i, i' = 1, 2, \dots, n)$$

(2)不偏性の仮定：

$$\varepsilon_i \text{ の期待値はゼロである。} \quad E[\varepsilon_i] = 0 \quad (i = 1, 2, \dots, n)$$

(3)等分散性の仮定：

$$\varepsilon_i \text{ の分散は全て等しい。} \quad V[\varepsilon_i] = \sigma^2 \quad (i = 1, 2, \dots, n)$$

(4)正規性の仮定：

$$\varepsilon_i \quad (i = 1, 2, \dots, n) \text{ は正規分布に従う。}$$

更に目的変数が  $q$  個ある場合、  
共分散に関して、

$\text{Cov}[\varepsilon_{ij}, \varepsilon_{ij}] = \sigma_{jj} \neq 0$  ( $i = 1, 2, \dots, n$ ,  $j, j' = 1, 2, \dots, q$ ) となる等分散共分散性の仮定がある。

多変量重回帰分析は  $q$  変量に相関が入っている為、変数全体に対して最尤法を用いる。その為、精度的には1変量の方が優れており、この事例集では1変量の重回帰分析を対象とする。

以下、個別の細かな前提条件は省略するが、他の解析手法もこの考え方と同じ様な基礎理論を土台にしている。

### 3. 事例

鋼鉄の生産管理に関して事例を紹介したい。

鉄鉱石から鉄を作る上で、その製造過程のデータを採取し、その鋼質の性質を多変量解析を用いて分析する。それまで経験上概略は分かっていたが不透明であった幾つもの事柄を、得られた観測データから分析し、その結果を考察する。各解析手法は一般的な方法で求め、どのような形で考え方を進めて行くかを論述する。

#### 3-1. 説明文

鉄鉱石はコークスや石灰石と共に溶鉱炉で製鉄を終えた後、トビードカによって製鋼工場へ運ばれる。製鋼工場では湯鍋に移された後、転炉へ入れられる。この転炉では吹錬を終えた後、取鍋に移される。転炉から出る温度が出鋼温度であり、ここでCとMnの値が測定される。転炉から取鍋に移される時、あらかじめ定められているCとMnの目標値に対してその差を埋め合わせるために合金であるHC(ルカーボンFe-Mn)、Si-Mn、Fe-Si、合金Alが投入される。

ついで鑄型に鑄込む作業時に採集された値がレドール成分である。この鑄型から抜き取ったものが鋼塊と称し、分塊工場で圧延処理が行われる。この時の半製品がスラブと呼ばれこのスラブ鋼を厚板工場で、鋼塊と厚板等の圧延処理が行われる。ここでフェック成分の値が採集される。ここで厚板製品を一部切り取り顕微鏡で格子に当たる確率から線状のゴミ、点状のゴミを求め清浄度(A系、B系)として、測定される。この清浄度が鋼質の値に影響を及ぼしている事が経験上分かっている。金属学の理論的立場から、レドール成分Sの影響があると考えられている。その為、Sの影響を除いた補正系等も求める(C系)。機械試験として、製品の引っ張り試験時の破面の状態を三段階の判定(良(1)、可(2)、不可(3))を行い、機械試験値(Y.P.降伏点、T.S.抗張力、E1伸び等)も測定する。

#### 3-2. 目標ターゲット

清浄度の値が高いと鋼質が悪化する事が、経験上からも知られている。これはどの工程でどの程度影響されているのか、機械試験値と鋼質の関係から、良好な鋼質を作るにはどうすれば良いのかを研究する。

#### 3-3. データ

チャージ NO	材質	板厚	圧延比			判定	フェック成分				
			鋼塊 / 製品	スラブ / 製品	鋼塊 / スラブ		C	Si	Mn	P	S
K12773	AF102R	6	133.8	25.0	5.35	1	14	22	39	13	15
K12774	AF102R	6	133.8	16.7	8.01	1	13	21	47	26	13
K12775	AF102R	6	138.8	16.7	8.01	1	12	21	49	9	16
K12776	AF102R	40	21.5	4.5	4.78	1	14	26	69	10	14
K12777	AH150B	9	84.7	11.1	7.63	1	15	30	75	12	15
.	.	.	.	.	.	.	.	.	.	.	.

フィック成分			レドール成分						終点成分		レドール差		投入合金属			
Al	N	O2	C	Si	Mn	P	S	Al	C	Mn	C	Mn	Fe-Mn	Si-Mn	Fe-Si	Al
43	45	16	13	23	41	13	14	36	8	29	6	12	860	0	940	240
14	51	24	13	25	49	17	13	25	10	22	3	27	1030	0	920	214
30	55	11	11	23	49	13	13	39	6	20	5	29	1140	0	930	297
32	60	19	15	25	70	10	14	25	6	20	9	50	2260	0	930	300
59	90	9	16	29	75	14	15	40	5	17	11	58	2750	0	910	270
.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.

温度		鋳込時間 min	鋳型	清浄度				機械的性質		
出鋼温度	鍋中温度			A系	B系	C系	合計	Y.P.	T.S.	El
1630	1600	12.8	u7	38	4	28	42	30.0	42.5	28.9
1640	1600	21.0	u7	33	13	43	46	32.7	48.3	21.3
1650	1625	15.20	u7	67	0	47	67	34.3	44.8	31.9
1645	1610	15.6	u5	67	8	67	75	26.4	43.3	27.1
1630	1600	15.7	u6	83	4	73	87	33.0	47.3	26.5
.	.	.	.	.	.	.	.	.	.	.

(表1)

(得られたサンプル数 n=120 個)

#### 4. 解析手法適用の基本的考え

主たる考え方は次の様なものである。

(但し、頁数の制限から一部の手法は事例集に記載していないものもある。)

考え方の手順として、このデータ解析において、分析の手順としては標準統計量、平均値の差の有意差検定 (T検定) 等により、その後の多変量に当てはめられ易くするため、要因毎の影響を想定して、データを分類し、変数毎のヒストグラム等を調査し、標準統計による2~3の変数間の有意差検定、各変数の影響の有無としての相関関係 (X-Yプロット、散布図) の調査、分散分析による影響度の判断、傾向値から予想した重回帰分析 (説明変数から目的変数への影響度の把握)、主成分分析、因子分析等による固有値、固有ベクトルの計算、更に主成分や共通因子の抽出からどの変数が強く影響を及ぼすものかを調査、目的とされる機械的性質が途中での清浄度により、全体としてどの様に作用を及ぼしているかを調査し、判別分析による母集団の判定を考慮し、更に多変数を全体としてとらえて、正準分析による変数の影響力と分類されたグループ間とのバランス、正準相関分析による2変数群間の調査も行い、各有効性も検討する。

これらを箇条書きにすれば以下の様なものである。

生産工程のラインを把握し、どの工程でどのデータが得られ、それが次の工程にどう影響し合うか、製品の品質の向上とはどれを目指すか等の目標を明確にする。生産管理全体の理解が進む。

( 生産工程の図示は省略する。 )

品質の向上から、解析対象とするデータを適切に分類する。原因を探る為、データ値の大小の形が後々理解しやすくこの時、サンプル上は等分類が望ましい。

( 分類のデータ表示は省き、T検定時に分類されたデータの一部の平均値のみ記載する。 )

各変数のヒストグラムを表示し、分布を確認する。分布の状態が理解しやすい。

( 事例集に記載せず。 )

目的とする要件に対して変数間の概略影響を調査する為、対象とする変数間の2分類、3分類に標準統計量による平均値の差の有意差検定 ( T検定 ) を実施する。全体的傾向値を把握するのに望ましい。

各変数間の相関分析を行い、全変数間相互の傾向値を調査する。これは単一の変数間同士で理解がし易い。

( 事例集に記載せず。 )

#### [解析の種類]

変数間の影響が目的とする要件に対して有意性が認められるか、否かを知る為、必要変数を選び分散分析を行う。

要因の影響を表現していると考えられる式をたて、その変数のもたらす寄与を検討する。重回帰分析 ( 段階式 ) を用いる。目的変数と説明変数の関係や重相関係数等を理解する。影響が大きく左右している変数群の全分散を考え、状況を総合的に把握する為、主成分分析を適用する。主成分の得点から主成分平面上に最大に分かれたグループから、もとのサンプルの変数値を抽出し、比較検討する。

各変数の影響はどの変数とどの変数が似ているか等の調査に因子分析を適用し、共通の因子を導きその因子軸の意味合いを調査する。

一部の製品がどの製品群からのものであるかを調査する為、判別分析を適用する。

更に、品質の向上にはどの変数をアクションとすべきかも合わせて検討する。

目的とする変数を良、可という形で分類し、それに対応した変数群を選び、どの影響が大きいかを調査する上で、正準分析を適用する。

( 解法上幾つかの考えはあるが、ここでは Bartlett 理論に基礎をおく H.L.シルの展開式を活用する、この項は一部分の数式記載。 )

2 変数群間の影響度を調査する上で、考えられる変数群を比較検討する為、正準相関分析を用いる。変数間同士の全体的傾向、及び正準軸を考えることにより、影響度が理解出来る。

以上、製品の品質向上を目指して、これらの解析結果から総合的にどのようなアクションを講じれば、良いかを検討する。

## 5. データの分類

鋼質の機械的性質は清浄度に影響を及ぼされていることが、経験的に分かっている。この清浄度に影響を及ぼすのはどの要因であるのか。データを多変量解析で解析する前に全体の傾向値を把握する為、データを幾つかに分類分けを行う。この時の分類方法として、全観測値を出来るだけ等分割する。標準統計量、正準分析を対象とした場合は2分類、又は3分類に、分散分析を目的とした場合は5分類にする。

例えば、清浄度全体においては平均値に近い点で良と不可に2分類し、これに対応する説明変数の平均値の差が標準統計量のT検定で有意差があるか否かの検定を行う。正準分析に対しては、清浄度A系において測定値に低いグループから良(Ga)、可(Gb)、不可(Gc)と3分類し、これに対応する説明変数を対象とする事により、この良や可の変化が何によって生じているのかを明らかにする。

こうした考え方でデータを適宜分類して行く。

## 6. 標準統計量の導出と検定 (T検定)

2つの変数間の平均値の差の有意性をT検定により行う。分類分けされた目的変数(清浄度等)が、様々な要因によってその区分が生じていると考え、説明変数の要因の一つ一つが目的変数の差を作っていると仮定した時、この仮説が正しいか否かのおおざっぱな判断を行う。つまり分類分けされた目的変数に対する一要因毎に、各グループ間の平均値の差の有意性をT検定により行う。

t分布の密度関数 (nは自然数)

$$f(x) = \frac{1}{\sqrt{n\pi}} \frac{\Gamma\left(\frac{n+1}{2}\right)}{\Gamma\left(\frac{n}{2}\right)} \frac{1}{\left(1 + \frac{x^2}{n}\right)^{\frac{n+1}{2}}} \quad (-\infty < x < \infty)$$

t分布の限界点

危険率 d.f.	5 %	1 %
30	2.04	2.75
40	2.02	2.70
60	2.00	2.66

分類されたデータから目的変数に対応する各要因のグループ間の平均値

	清浄度 A 系			清浄度全体		鋼質判定	
	良	可	不可	良	不可	判定 1(良)	判定 3(不可)
板厚	16.25	18.33	16.00	16.56	17.22	16.76	16.33
レドール成分 C	16.05	16.11	14.71	15.64	15.03	14.85	15.51
"  Si	25.95	32.94	29.57	27.92	30.66	28.19	29.59
"  Mn	87.00	113.22	97.57	92.20	105.10	88.57	104.50
"  P	15.65	18.00	16.85	16.10	17.37	16.42	17.03
"  S	12.01	13.33	17.10	12.32	15.18	13.71	13.63
"  Al	33.60	33.77	31.35	33.24	32.88	34.33	32.59
出鋼温度	1645.20	1649.72	1644.53	1645.24	1647.81	1644.22	1647.71
.	.	.	.	.	.	.	.

(表 2)

目的変数に対する各要因のグループ間の平均値の差の T 値

\*印 5%有意、 \*\*印 1%有意

	清浄度 A 系			判定 1 良, 3 不可	清浄度 A 系と 鋼質判定	
	良&可	可&不可	良&不可	判 1& 判 3	清良&判 1	清良&判 3
板厚	0.71	0.63	0.07	0.16	0.07	0.09
レドール成分 C	1.23	1.82	0.36	0.88	1.06	0.15
"  Si	3.40**	1.20	1.88	0.69	0.15	0.83
"  Mn	2.94**	1.50	1.07	1.91	0.40	1.53
"  P	2.27*	1.03	1.04	0.62	0.25	0.93
"  S	2.05	5.69**	8.04**	0.10	2.38*	1.82
"  Al	0.05	0.74	0.56	0.56	0.34	0.20
出鋼温度	1.12	1.14	0.18	0.95	0.25	0.71
.	.	.	.	.	.	.

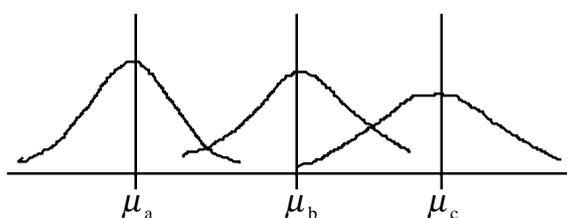
(表 3)

< 検討 >

清浄度 A 系に対して T 検定の結果、有意差が認められた説明変数はレドール成分 Si、P、S、Mn、及びレドール成分との差 Mn であった。

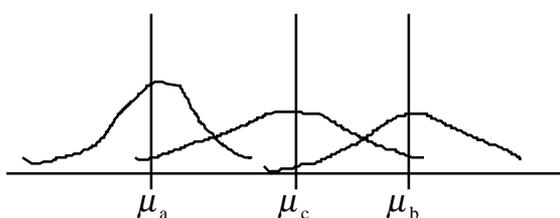
清浄度 A 系に対する Ga (良)、Gb (可)、Gc (不可) と 3 分類すると、これに対応する平均値をそれぞれ  $\mu_a$  ,  $\mu_b$  ,  $\mu_c$  とおけば、 $\mu_a < \mu_b < \mu_c$  又は

$\mu_a > \mu_b > \mu_c$  の何れかになれば、この説明変数は清浄度 A 系に影響を及ぼしていると考えて良い。(図 1 参照)



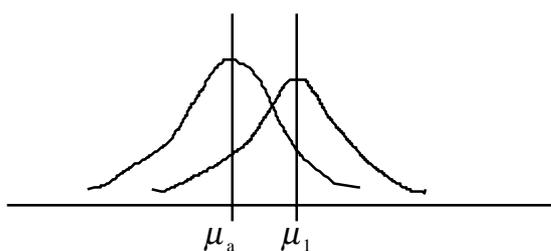
(図 1)

清浄度 A 系の 3 分類に対するレドール成分 S の分布



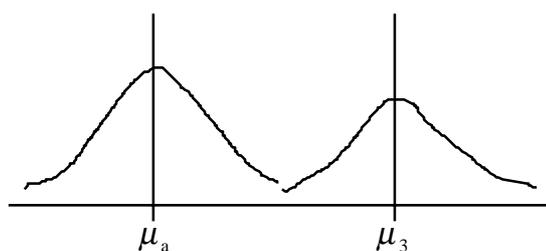
(図 2)

静浄度 A 系の 3 分類に対するレドール成分 Si の分布



(図 3)

対応する二つの各要因間の平均値に有意差がない場合



(図 4)

対応する二つの各要因間の平均値に有意差がある場合

#### 有意差検定 (T 検定) に対する影響力の考え方

清浄度 A 系に対応するレドール成分 S の理論分布 (図 1) とレドール成分 Si の分布 (図 2) の 2 つの場合、(図 1) の方はそれぞれの平均値が  $\mu_a < \mu_b < \mu_c$  又は  $\mu_a > \mu_b > \mu_c$  (第一パターン) となってそれぞれ有意差があるが、レドール成分 Si に対してはそれぞれの平均値が  $\mu_a < \mu_c < \mu_b$  (第二パターン) となっており、 $\mu_a$  と  $\mu_b$  には有意差が認められるが、第二パターンにおける差は Si だけの影響力ではなく、他の要因の影響があると推定される。つまり第二パターンは  $\mu_a$  と  $\mu_b$  には他の影響力も含めていろいろ偶然の差が入り交じったものが考えられると判断したい。

つまり第一のパターンに適していない場合は、その説明変数が目的変数に対してグループ分けに影響を及ぼしていないと判断しても良いと考えられる。

浄度全体が鋼質の判定に影響を及ぼしているか否かについて、以下の様に考えたい。清浄度全体が鋼質の判定に影響を及ぼしている要因が存在するならば、清浄度全体良（Ga）に対する説明変数と判定1（良）に対応する説明変数間に有意差がなく（図3参照）、判定3（不可）に対応する説明変数間に有意差が認められる場合である。（図4参照）かつ清浄度全体不可（Gc）に対応する説明変数と判定3（不可）に対応する説明変数に有意差がなく、判定1（良）に対応する説明変数に差がある場合である。ここで清浄度A系のGa（良）とこれに対応する鋼質判定1（良）、3（不可）のデータとの関係では各要因により、有意差がありなしの理論分布を示す。（図3、図4参照）

#### <考察>

標準統計量における平均値の差の有意差検定（T検定）で判明出来たものを述べる。

- (1)清浄度A系に関しては、レドール成分Sが影響していると推測出来る。
- (2)清浄度A系に関しては、レドール成分Si、Mn、Pが少し位ではあるが、何らかの影響を及ぼしている可能性が考えられる。
- (3)清浄度B系に関しては合金Fe-Siが影響していると推測出来る。
- (4)清浄度全体では、レドール成分S、及び合金HC+SiMnが主たる影響を及ぼしている。
- (5)フィック成分O<sub>2</sub>は、終点C及びレドール成分との差Cの影響力が強い。
- (6)清浄度A系全般は、引っ張り試験時の判定に対して、余り影響を及ぼしていないと推測出来る。
- (7)レドール成分Sが引っ張り試験時の判定に良くない影響が考えられる。
- (8)出鋼温度、鑄込時間、圧延比は清浄度A系B系に対しても、引っ張り試験時の判定に対しても、余り影響を及ぼしていないと考えられる。
- (9)判定1（良）と判定3（不可）の影響区分は、合金HC及びFe-Siであると考えられる。
- (10)清浄度A系とC系の相違は、レドール成分Sの影響力が考えられる。
- (11)清浄度全体に影響を及ぼしている要因は、合金HC+SiMn（ルカーボンFe-MnとSiMnの合計値）であると判断できる。

7 . 解析手法

この章では多変量解析の代表的な手法を使って、鋼質の品質分析を行う。

7 - 1 . 分散分析(Analysis of Variance)

データ全体のバラツキを平方和で表現し、これを処理効果を反映する平方和と、誤差のみ反映する平方和に分解して、その時の処理効果と誤差を比較する。2つの分散の検定である為、F分布を用いた検定を行う。

< 解析目的 >

出鋼温度、鍋中温度、鋳込時間、鋳型、圧延比率等の個々の測定値の変化が清浄度に対して影響を及ぼしているか否かの一つの判断尺度として、分散分析一元配置を用いる。

< 解析内容 >

温度、時間、圧延比等の分類基準は既述のデータ分類方法による。

F分布の密度関数 (m,n は自然数)

$$f(x) = \frac{\Gamma\left(\frac{m+n}{2}\right) m^{\frac{m}{2}} n^{\frac{n}{2}}}{\Gamma\left(\frac{m}{2}\right)\Gamma\left(\frac{n}{2}\right)} \frac{x^{\frac{m}{2}-1}}{(mx+n)^{\frac{m+n}{2}}} \quad (0 < x < \infty)$$

( x = 0 の時 f(x)=0 )

F分布の限界点

d.f. n \ d.f. m	3		4	
	5%	1%	5%	1%
30	2.92	4.51	2.69	4.02
40	2.84	4.31	2.61	3.83
60	2.76	4.13	2.53	3.65

< 計算結果 > 分散分析一元配置

目的変数	対象とした変数				
		出鋼温度	鍋中温度	鋳込時間	鋳型
清浄度全体	F値	0.29	1.13	1.26	1.18
	有意性	×	×	×	×

目的変数	対象とした変数				
		圧延比(鋼塊/スラブ)	圧延比(スラブ/製品)	終点 C	終点 Mn
清浄度全体	F 値	0.53	0.97	1.33	1.80
	有意性	×	×	×	×

目的変数	対象とした変数						
		出鋼温度	鍋中温度	鋳込時間	鋳型	終点 C	終点 Mn
フィック成分O <sub>2</sub>	F 値	0.27	1.45	1.53	1.09	0.65	0.98
	有意性	×	×	×	×	×	×

< 考察 >

分散分析での計算結果は上記の如く、対象した変数がいずれの場合も有意性は認められなかった。ここで×印で示す。これから判断出来る事は出鋼温度 鍋中温度 鋳込時間 鋳型圧延比率等の測定値の変化は清浄度全体の合計や、フィック成分O<sub>2</sub>には影響を及ぼしていないと判断することが出来る。

(注) 分散分析そのものは、多変量解析の一分野として扱われていない書籍が一般的であるが、ここでは事例集における考え方のステップから1節を設けた。

## 7 - 2 . 重回帰分析(Multiple Regression Analysis)

複数個の説明変数と1つの目的変数の関係式を推定する。この関係の値を重相関係数(multiple correlation)と呼ぶ。正規方程式を導き最小二乗法により求められる一般型重回帰分析を始め、総当たり法、変数増加、減少、増減、減増法や変数指定選択結合法等がある。ここでは、偏F検定による変数増減法を用いた段階式重回帰分析による最良な回帰式を導く。

< 解析目的 >

清浄度(A系、B系、C系、全体合計)、フィック成分O<sub>2</sub>及び機械的性質(Y.P.,T.S.,EI)に対して、段階式重回帰分析による説明変数による選択と、その時の重回帰式の意味合いを検討する。次の説明変数(独立変数 independent variable)のすべてを対象として取り上げ、変数増減法で計算した時、採択された変数による重回帰式を掲載する。目的変数(従属変数 dependent variable)である清浄度に対してはレドール成分群を、機械的性質に対してはフィック成分を対象とする。ここでフィック成分O<sub>2</sub>は鋼質の生産管理では、その工程上目的変数として取り扱われる事がある。いずれの説明変数に対しても、出鋼温度、鍋中温度、鋳込時間、圧延比(鋼塊/製品)及び合金成分(HC+SiMn(フェリウムとシリウムの合計値)、Fe-Si、Al)は同じく取り入れる。

< 計算結果 >

( )内は回帰係数の t 値、 [ ]内は目的変数 (従属変数) Y との単相関係数、 R は重相関係数、  $\sigma^2$  は推定値の標準誤差

Y : 清浄度 A 系、  $x_1$  : レド成分 S、  $x_2$  : 合金 HC+SiMn

$$Y = -42.870 + 7.934 x_1 + 0.005 x_2, \quad R=0.694, \quad \sigma^2 = 24.456$$

(6.35)      (2.25)  
[0.65]      [0.23]

Y : 清浄度 B 系、  $x_1$  : 合金 Fe-Si

$$Y = 17.577 - 0.011 x_1, \quad R=0.392, \quad \sigma^2 = 7.988$$

(- 3.01)  
[ - 0.39]

Y : 清浄度 C 系、  $x_1$  : 合金 Fe-Si

$$Y = 110.149 - 0.035 x_1, \quad R=0.348, \quad \sigma^2 = 27.381$$

(- 2.62)  
[ - 0.34]

Y : 清浄度合計、  $x_1$  : レド成分 S、  $x_2$  : 合金 HC+SiMn

$$Y = -41.333 + 8.140 x_1 + 0.007 x_2, \quad R=0.722, \quad \sigma^2 = 24.465$$

(6.52)      (3.25)  
[0.65]      [0.32]

Y : 引く成分 O<sub>2</sub>、  $x_1$  : 鋳込時間、  $x_2$  : 圧延比 (鋼塊 / 製品)

$$Y = 33.680 - 0.622 x_1 - 0.059 x_2, \quad R=0.416, \quad \sigma^2 = 6.820$$

(- 2.51)      (- 2.05)  
[ - 0.31]      [ - 0.25]

Y : Y.P.降伏点、  $x_1$  : 引く成分 Mn、  $x_2$  : 合金 Fe-Si、  $x_3$  : 圧延比 (鋼塊 / 製品)

$$Y = 11.735 + 0.148 x_1 + 0.004 x_2 + 0.083 x_3, \quad R=0.793, \quad \sigma^2 = 2.867$$

(8.09)      (2.64)      (6.13)  
[0.56]      [ - 0.23]      [0.24]

Y : T.S.抗張力、  $x_1$  : 引く成分 C、  $x_2$  : 引く成分 Si、  $x_3$  : 引く成分 Mn、

$x_4$  : 引く成分 N、  $x_5$  : 合金 Fe-Si、  $x_6$  : 圧延比 (鋼塊 / 製品)

$$Y = 19.393 + 0.591 x_1 + 0.008 x_2 + 0.159 x_3 - 0.049 x_4 + 0.006 x_5 + 0.059 x_6$$

(4.90)	(0.10)	(6.40)	(- 2.09)	(3.58)	(6.27)
[0.66]	[0.71]	[0.69]	[0.14]	[- 0.11]	[0.004]

, R=0.926 ,  $\sigma^2 = 1.962$

Y : El 伸び、 $x_1$  : チック成分 Si、 $x_2$  : チック成分 Mn、 $x_3$  : チック成分 S

$$Y = 40.884 - 0.379 x_1 + 0.057 x_2 - 0.535 x_3$$

(- 3.80)	(2.52)	(- 3.34)
[- 0.36]	[0.01]	[- 0.50]

, R=0.654 ,  $\sigma^2 = 3.219$

#### <考察>

清浄度 A 系はレドール成分 S と合金 HC+SiMn が採択されたが、S の影響が強く、S によって左右されている事が分かる。清浄度 A 系を良くする為には S を一定限界内で、値を低くする事が望ましい。清浄度 B 系、C 系は合金 Fe-Si の値を高く設定する様にする方がよい。清浄度の合計はレドール成分 S 以外に合金 HC+SiMn に影響される事であり、この合金の値は下げる方がよい。

チック成分 O<sub>2</sub> の値を低くするには現状以上に圧延比（鋼塊 / 製品）を大きくし、鑄込時間を長くする方向が良いと考えられる。

降伏点 Y.P. においてはチック成分 Mn の影響力が大きく、この値は上げる方が望ましい。抗張力 T.S. においてはチック成分 C、Mn の影響力が大きく、いずれもこの値は上げる方が望ましい。伸び El はチック成分 S 及び Si に影響が及ぼされているものと考えられ、いずれも下げる方が望ましいと判断出来る。

つまり機械的性質全体を良くする為には、チック成分 C、Mn の値は上げる方が良く、S、Si の値は下げる方がよいと判断できる。どの程度の値まで上げたり下げたりすべきかは、他の多変量解析から範囲を推定したい。

### 7 - 3 . 主成分分析(Principal Component Analysis)

p ( 2 ) 個の変数の特性値の持つ情報を、より少ない m 個 ( m < p ) の総合特性値に要約する手法である。空間内で分散が最大となる軸 ( 主成分 ) を必要な固有値分次々と見つける。そして m 個の主成分は互いに無相関である。場合により座標軸の回転も考えられている。

#### <解析目的>

鋼質の性質を決定すると考えられているチェック成分の値から、2 ~ 3 の主成分を導出し、主成分平面にサンプルをプロットした所、主成分のある位置にそれぞれのサンプルがプロットされる所から、その時の観測値を抽出する事によって、異なるグループ間の機械的性質と、チェック成分の値を比較検討し、良質な機械的性質をもたらすチェック成分の

値の範囲をもとめる。但し、鋼質の性質から O<sub>2</sub> (酸素) は外す。

< 計算結果 >

固有値寄与率結果

固有値番号	固有値	寄与率	累積寄与率
1	976.31	0.731	0.731
2	180.93	0.135	0.866
3	117.45	0.087	0.954
4	26.22	0.019	0.974
5	21.79	0.016	0.990
6	7.34	0.005	0.995
7	5.45	0.004	1.000

主成分軸

第 1 主成分軸

$$z_1 = 0.0332C + 0.1625Si + 0.9622Mn + 0.0518P - 0.0116S + 0.1644Al + 0.1295N$$

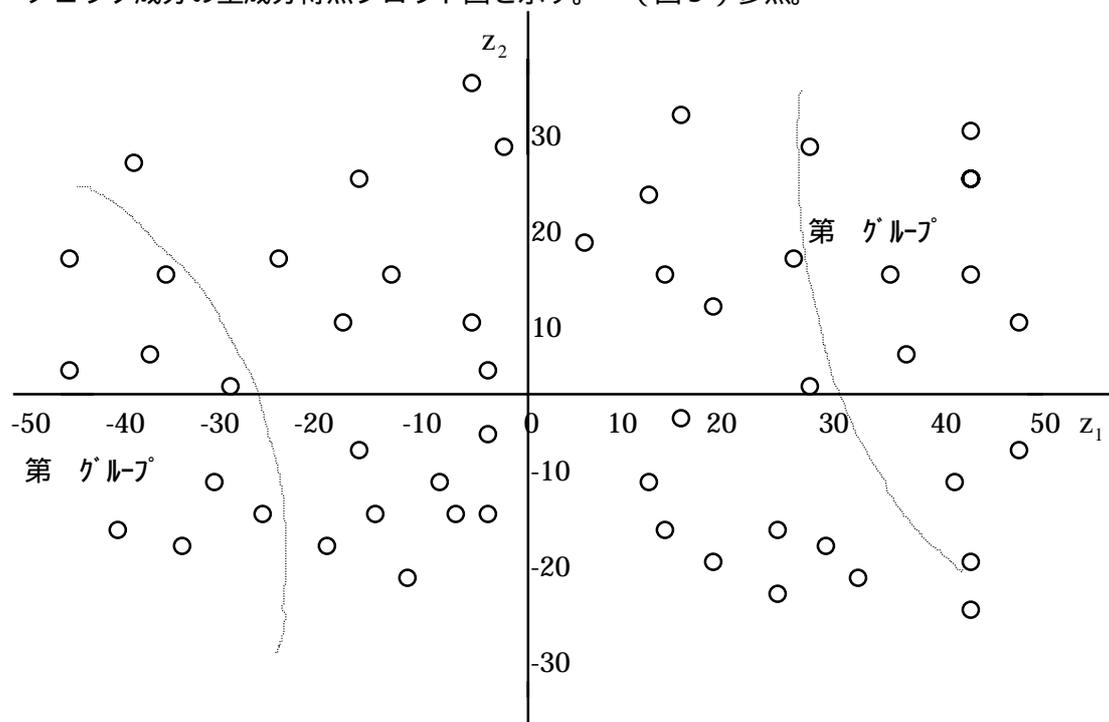
第 2 主成分軸

$$z_2 = 0.0270C - 0.0085Si - 0.1680Mn - 0.0987P + 0.0275S + 0.9786Al + 0.0517N$$

第 3 主成分軸

$$z_3 = 0.0071C + 0.0224Si - 0.1209Mn - 0.0529P - 0.0089S - 0.0782Al + 0.9878N$$

チェック成分の主成分得点プロット図を示す。(図 5) 参照。



(図 5)

(図5)より、主成分平面の2つの軸 $z_1$ 、 $z_2$ に対して、この固有値の寄与率が $z_1$ 対 $z_2$ で1対0.18、第一固有値の累積寄与率は73.1%となっている。

そこで主成分平面を主成分軸 $z_1$ の高い所で、かつ $z_2$ のやや高めのグループを第1グループ、主成分軸 $z_1$ の低い所で、かつ $z_2$ のやや低めのグループ(第1グループと正反対のグループ)を第2グループとする。

これは主成分分析で各変数の群がもっとも離れた所に位置する為に、目的とする変数に及ぼしている影響が反対のものであることが予想できる。

つまりこの2グループを比較する事によって、目的とする変数(今回は機械的性質)の向上を目指すことが予想出来る。

主成分平面のプロット図から抽出した比較表

サンプル 番号	第1グループ									
	機械的性質			チェック成分						
	Y.P.	T.S.	El	C	Si	Mn	P	S	Al	N
34	39.8	49.8	31.0	14	28	72	19	13	33	56
35	37.6	54.3	29.4	16	27	73	20	13	32	55
40	38.9	55.6	26.8	16	30	75	17	12	32	53
44	36.4	53.8	30.9	17	32	75	17	12	32	53
・	・	・	・	・	・	・	・	・	・	・
平均値	38.2	53.9	29.3	16.3	30.8	74.6	18.1	12.9	32.3	54.3

(表4)

サンプル 番号	第2グループ									
	機械的性質			チェック成分						
	Y.P.	T.S.	El	C	Si	Mn	P	S	Al	N
26	28.3	42.3	23.5	14	25	59	24	16	39	56
29	31.5	40.6	21.6	13	25	52	21	15	42	55
52	30.6	41.9	24.6	12	22	53	21	15	41	54
55	29.8	39.9	21.9	15	24	54	20	14	40	55
・	・	・	・	・	・	・	・	・	・	・
平均値	30.5	41.5	22.4	14.1	23.8	55.6	22.0	15.8	40.5	55.8

(表5)

#### <考察>

鋼質の機械的性質を決定していると考えられるチェック成分の主成分分析結果は上記

の形であった。第一主成分の式  $z_1$  より、各説明変数に付加している固有ベクトルに着目すれば Mn が圧倒的に大きい。これはこの変数の分散が最大である事からも予想出来る。第二主成分の式  $z_2$  では Al の固有ベクトルが最も大きい。つまり第二主成分では Al の影響力があるものと考えられる。

主成分得点プロット図 (図 5) で主成分の軸から考えると、 $z_1$  軸の高い所に位置しているサンプルは Mn の影響力が大きく、 $z_2$  軸の高い所に位置しているサンプルは Al の影響力が大きいものと判断できる。

その時のウェイト比は約 7 対 1 である。他の解析より Mn の影響が大きいのは機械的性質のうち Y.P. 及び T.S. であり、 $z_1$  軸の高い所に位置しているサンプルは機械的性質の重要な部分であろうと予想出来る。

ここで鋼質においては、機械的性質の Y.P. と T.S. は弱い正の相関があるが、Y.P. 又は T.S. と El とは相関関係が低い事が知られている。従来の実験ではこの 3 つの値は相反するもので、一方を上げても他方の値は必ずしも上がらず、全体を良くする事は困難と考えられて来た。

主成分分析の活用により、全体の値を向上させる事を考える。第一主成分軸及び第二主成分軸 (固有値比率から相反する度合いの部分) の最大分散で、最も離れた 2 つのグループ (表 4) と正反対の第二グループ (表 5) に位置する主成分得点からプロットされるものからサンプルを取り上げ、チェック成分の平均値及び標準偏差を求める。これらから目的とする機械的性質の向上が期待される。

機械的性質 (Y.P., T.S., El) を向上させる為のチェック成分の最適値は以下の範囲になる。

C :  $16.3 \pm 1.2$  , Si :  $30.8 \pm 2.1$  , Mn :  $74.6 \pm 2.6$  ,  
P :  $18.1 \pm 1.5$  , S :  $12.9 \pm 1.6$  , Al :  $32.3 \pm 2.0$  ,  
N :  $54.3 \pm 2.4$

#### 7 - 4 . 因子分析 (Factor Analysis)

分析の対象となる多変量間の相関は、各変量に潜在的に共通に含まれている少数個の共通の因子により生じる考えからなる。因子軸の係数は因子負荷量 (factor loading) として求められる。共通因子 (common factor) の意味解釈の為、因子軸を回転させて解釈する事が多い。空間内において各因子軸は必要な固有値分求められ、それぞれ互いに無相関になる。

< 解析目的 >

レドール成分群、鍋中温度、鋳込時間、圧延比、合金属群等 鋼質を決定している様々な要因の中から共通な因子を抽出し、その共通因子の意味を考えると共に、因子負荷量の最尤推定量を求め、鋼質に影響力をもたらす変数間の共通の意味を吟味する。

< 計算結果 >

累積寄与率が70%に達するまでの5番目迄の固有値を求める。

固有値番号	固有値	寄与率	累積寄与率
1	4.19	0.32	0.32
2	1.92	0.15	0.47
3	1.34	0.10	0.57
4	1.08	0.09	0.66
5	0.88	0.06	0.72

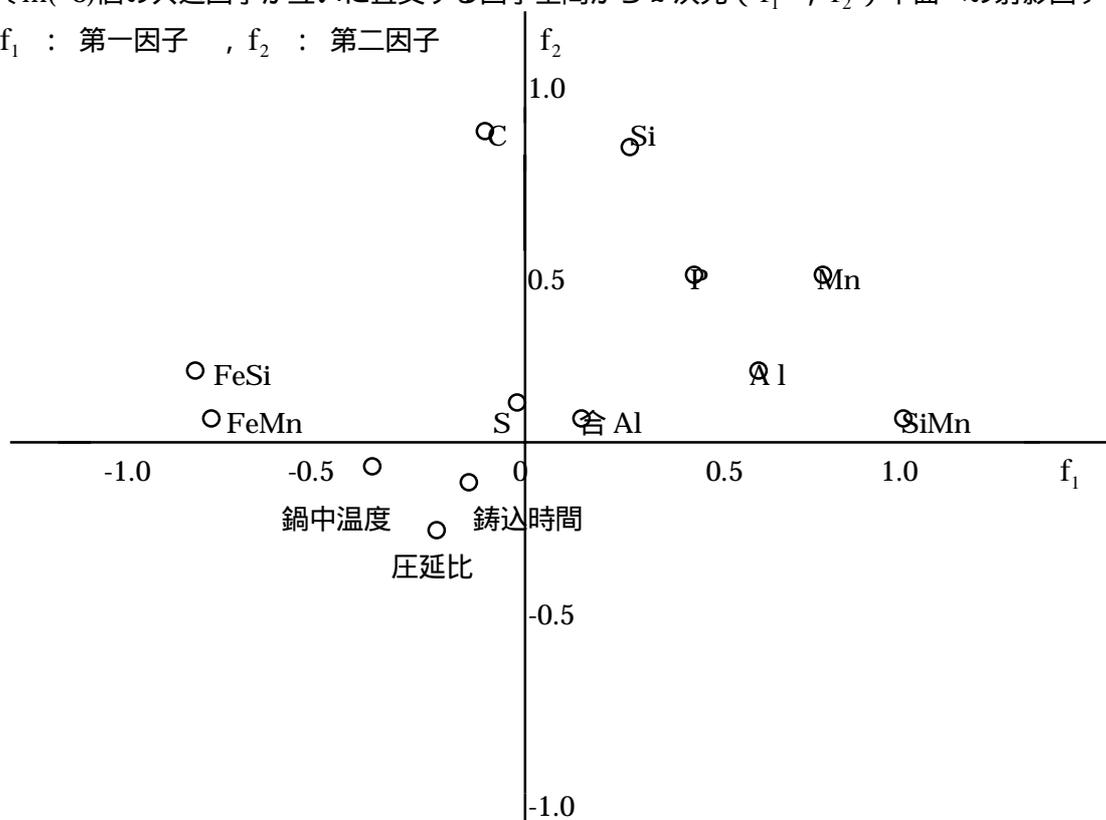
正規化された因子負荷量の最尤推定量 (第三因子迄)

変数名	第一因子	第二因子	第三因子
圧延比 (鋼塊/製品)	- 0.3587	- 0.3553	0.0657
レドール成分 C	0.2062	0.7039	0.2175
" Si	0.6868	0.5193	0.1635
" Mn	0.9090	0.1785	0.0334
" P	0.6106	0.3204	0.1152
" S	- 0.1530	0.0042	0.7864
" Al	0.5613	- 0.0756	0.1041
鍋中温度	- 0.3908	0.3547	- 0.1537
鋳込時間	- 0.2908	0.1158	0.4051
合金属 Fe-Mn	- 0.6665	0.4304	- 0.1544
" Si-Mn	0.8794	- 0.2630	0.1433
" Fe-Si	- 0.6722	- 0.2630	0.2013
" Al	0.3305	0.4087	- 0.6305

標準化リッジ法による回転後の因子負荷量の最尤推定量（第三因子迄）

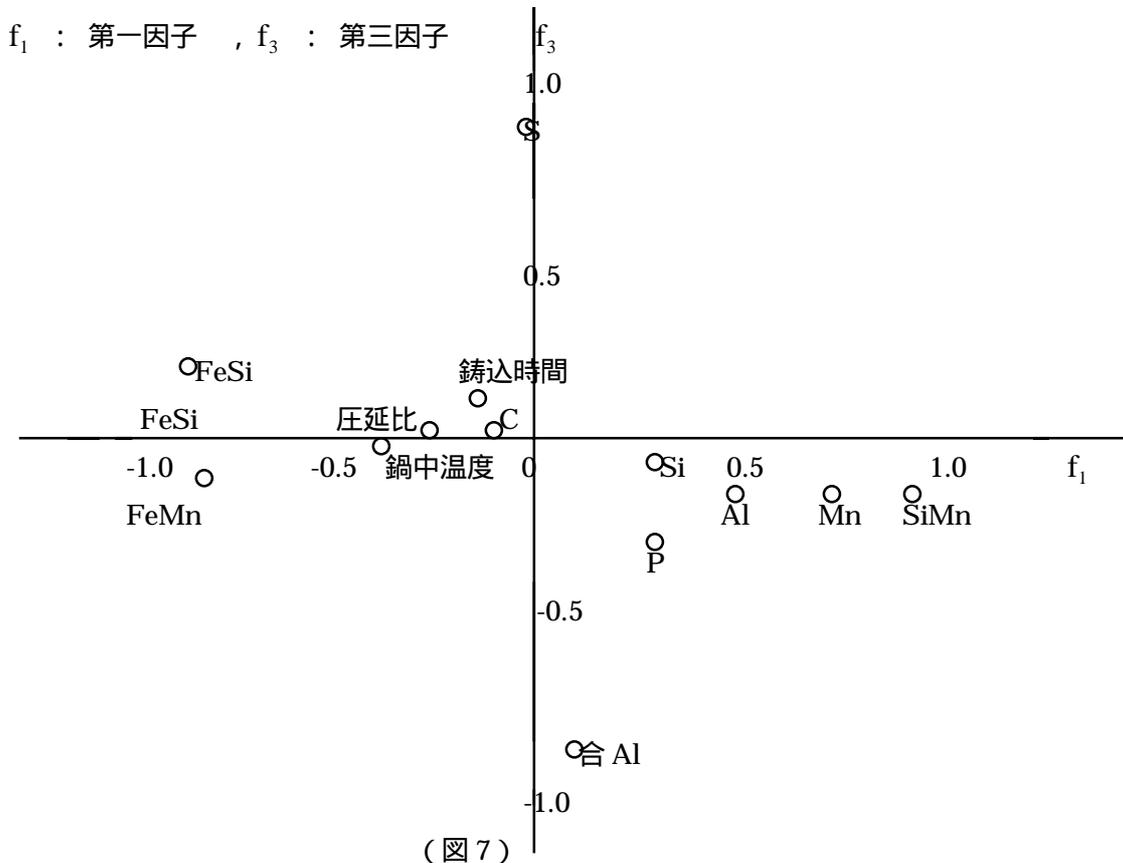
変数名	第一因子	第二因子	第三因子
圧延比（鋼塊/製品）	- 0.2436	- 0.2434	0.0960
レドール成分 C	- 0.1779	0.8242	0.0502
"  Si	0.3544	0.8007	- 0.0802
"  Mn	0.7149	0.5427	- 0.1946
"  P	0.3704	0.5320	- 0.3007
"  S	- 0.0179	0.1184	0.8039
"  Al	0.5153	0.2546	- 0.0158
鍋中温度	- 0.4083	- 0.1269	- 0.0055
鋳込時間	- 0.1890	- 0.1379	0.1934
合金属 Fe-Mn	- 0.8073	0.0145	- 0.1030
"  Si-Mn	0.9256	0.1702	- 0.0225
"  Fe-Si	- 0.8578	0.2706	0.2194
"  Al	0.0452	0.2317	- 0.7805

（図 6）は各変量  $x_i$  ( $i = 1, 2, \dots, p$ ) が共通因子 ( $f_1, f_2$ ) の張る空間へ射影した時の図  
 [  $m(=3)$ 個の共通因子が互いに直交する因子空間から 2次元 ( $f_1, f_2$ ) 平面への射影図 ]  
 $f_1$  : 第一因子 ,  $f_2$  : 第二因子



(図 6)

(図7)は各変量  $x_i$  ( $i = 1, 2, \dots, p$ ) が共通因子 ( $f_1, f_3$ ) の張る空間へ射影した時の図



<考察>

(図6)から判断できることは、第一因子軸はレドール成分 Mn、Al、合金 SiMn が位置している所から機械的性質の降伏点 Y.P.ないし抗張力 T.S.を表す因子である事が推定出来る。又合金 SiMn はレドール成分 Mn に近い性質をしている事も分かる。合金 FeSi、FeMn は機械的性質 (Y.P. , T.S.) を向上させるものではない事が想像出来る。同じ解釈のもとで、第二因子はレドール成分 C、Si が位置しているところから鋼質の硬さを表す因子と考えられる。第三因子はレドール成分 S が位置しており、鋼質の表面の美しさや滑らかさを表す因子と考えられる。合金 Al は鋼質の美しさに良い影響を与えているものではないと考えられる。2次元平面に存在するこれらのプロットから、各変数は一部のレドール成分群 (P, Al, Mn)、一部の合金群 (Fe-Si, Fe-Mn)、温度時間群 (圧延比、鍋中温度、鋳込時間) はそれぞれ近い所に位置され、似た様な性質があるものと考えられる。温度時間群は共通因子軸の原点近くに位置し、鋼質の機械的性質に余り影響を持っていないと解釈出来る。(図7)も合わせて考慮し、この因子分析結果からは、第三因子迄で表現出来る割合が全体の半分程度のものであった。

## 7 - 5 . 判別分析(Discriminant Analysis)

文献によっては判別関数(Discriminant Function)とも言う。新しいデータがそれ以前に得られたデータ群に対して、各データの群の重心(母平均)を求める事により、どの群に近いかを判別する。多次元空間上における距離の概念から成り立つ。これは異なる母集団が幾つかある時、この母集団の群を判別する関数を求める事にある。

### < 考え >

この事例では昨年得られたデータとの比較判断を始め、更にその判別される群を明確にさせる為、説明変数をどの様に動かせば、目的とする群(良の群)を更に良くさせる事が出来るかに関して検討してみたい。本来判別分析は、所属が分からない新しいサンプルが、それ以前に得られている情報から、いずれかの群に属するのか、あるいは全くどの群にも属しないと考える事が妥当であるのかを見極める事にある。この判別が難しいとき、変数を追加する事により、分離可能な関数を作るのが一般的である。ここでは目的とした群を更に向上させるにはどのようにすべきであるかの観点も加えて考えてみた。

### < 解析目的 >

いま昨年測定された生産工程改良前の取鍋時の値が幾つか存在している。この時期には、合金投入方法も少し異なっていた。この投入タイミングによる量の相違が清浄度のどちらの分野に位置づけられるかを判断してみたい。

(ケース1)の場合においての値

$x_1$  : 合金 Fe-Si 、  $x_2$  : 合金 HC+Si-Mn

昨年の合金投入の平均値

$$x_1 = 2450$$

$$x_2 = 940$$

### < 計算結果 >

(ケース1)

説明変数として 清浄度 A 系〔Ga(良)、Gb(可)、Gc(不可)〕対応する合金 Fe-Si,HC+Si-Mn の群を対象とする。

$x_1$  : 合金 Fe-Si 、  $x_2$  : 合金 HC+Si-Mn

$z_1$  : 2 群を判別する判別関数

清浄度 A 系(良 Ga)と(可 Gb)の判別関数

$$z_1 = -3.9724 - 0.0022 x_1 - 0.0007 x_2$$

清浄度 A 系(良 Ga)と(不可 Gc)の判別関数

$$z_1 = -3.5548 - 0.0022 x_1 - 0.0006 x_2$$

清浄度 A 系(可 Gb)と(不可 Gc)の判別関数

$$z_1 = 0.4176 + 0.0000001 x_1 + 0.0001 x_2$$

清浄度良のグループ Ga、清浄度可のグループ Gb とし、対象とするサンプルと群 Ga の母平均からの基準化された平方距離を  $D_a^2$  とし、群 Gb の母平均からの基準化された平方距離を  $D_b^2$  とする。（この  $D_a^2$ 、及び  $D_b^2$  のそれぞれの平方根がマハラビスの汎距離）  
計算結果

$$D_a^2 = 6.2 > D_b^2 = 2.7$$

この時  $D_b^2$  の方が少なく、対象とするサンプルは Gb グループに属することが分かる。つまり、昨年改良前の取鍋時の投入合金量の量は清浄度が良くならない方法であった、ものであると判断する事が出来る。

工程を改良した事により、投入タイミングと量の調整によって清浄度は昨年より良の方向の生産が期待できる。

次に 3 分類された清浄度 A 系に対して、それに対応する説明変数としてレドール成分 C、Mn の群、終点成分における C、Mn の群対象とした形で判別分析を行う。

< 計算結果 >

(ケース 2)

説明変数として 清浄度 A 系〔Ga (良)、Gb (可)、Gc (不可)〕に対応するレドール成分 C、Mn を対象とする。

$x_1$  : レドール成分 C 、  $x_2$  : レドール成分 Mn

$z_2$  : 2 群を判別する判別関数

清浄度 A 系 (良 Ga) と (可 Gb) の判別関数

$$z_2 = -4.4222 - 0.0850 x_1 - 0.0309 x_2$$

清浄度 A 系 (良 Ga) と (不可 Gc) の判別関数

$$z_2 = 0.0063 + 0.0947 x_1 - 0.0152 x_2$$

清浄度 A 系 (可 Gb) と (不可 Gc) の判別関数

$$z_2 = 4.4285 + 0.1798 x_1 + 0.0157 x_2$$

(ケース 3)

説明変数として 清浄度 A 系〔Ga (良)、Gb (可)、Gc (不可)〕に対応する終点成分 C、Mn を対象とする。

$x_1$  : 終点成分 C 、  $x_2$  : 終点成分 Mn

$z_3$  : 2 群を判別する判別関数

清浄度 A 系 (良 Ga) と (可 Gb) の判別関数

$$z_3 = 2.1828 + 0.0487 x_1 + 0.1140 x_2$$

清浄度 A 系 (良 Ga) と (不可 Gc) の判別関数

$$z_3 = 1.6000 + 0.2407 x_1 - 0.00005 x_2$$

清浄度 A 系 (可 Gb) と (不可 Gc)

の判別関数

$$z_3 = -0.5828 + 0.1919 x_1 - 0.1140 x_2$$

< 考察 >

(ケース 1) の場合、昨年得られたデータとの判別結果は、その工程を改良した事により、投入タイミングと量の調整によって清浄度は昨年より良くなっていると考えられる。投入すべき合金 Fe-Si, HC+Si-Mn は少な目の方が良いと推定出来る。

(ケース 2) (ケース 3) においては、判別関数の式からの解釈となる。

(ケース 2) の場合、3 つの判別関数の式からは、目的とした清浄度 A 系の善し悪しに関して線形分離可能ではない。3 つの判別関数は判別すべき事が明確には出来ていない。

つまり更に変数を増加させるか、あるいはレド成分 C, Mn だけが判別に寄与していないと考えられる。

(ケース 3) の場合、第二第三式より終点成分 C を上げ、終点成分 Mn は下げる様にすれば、2 つの群は明確になり、清浄度 A 系は良くなる方向へ向かうと考えられる。

## 7 - 6 . 正準分析(Canonical Analysis)

多次元空間内に正準軸を求め、複数個の変量母集団を対象とし、その母集団の平均値の回りにグループ毎に、正準空間内に信頼限界が球体として分布する中からそれぞれの意味合いを考える。空間内における正準軸は必要な固有値分求められ、それぞれの正準軸は互いに無相関になる。

[ Bartlett の基礎理論(1947)に基づく ]

< 概要 >

いま N 個の p 変量観測値が幾つか (q 個,  $q \geq 2$ ) ある時、このそれぞれの母集団を考える。第一母集団をベクトル  $\mu^{(1)}$  によって表される点を中心とし、この点の回りに分散共分散行列  $\sum^{(1)}$  によって特徴づけられるある球体の形をして、p 次元空間に分布していると考ええる。同じく第二母集団は p 次元空間に  $\mu^{(2)}$  及び  $\sum^{(2)}$  によってある球体の形で分布している。同様な考えで、残りの  $q - 2$  個の p 変量母集団に適用する。

但し、各母集団の点群の広がりはそれぞれ異なっている場合が多い。

仮に  $\mu^{(1)}$  と  $\mu^{(m)}$  ( $m \geq 2$ ) が一致するしないにかかわらず、 $\sum^{(1)} \neq \sum^{(m)}$  ならば、1 番目と m 番目の分布は異なったものになる。

ここで仮説の検定

$$\sum^{(1)} = \sum^{(2)} = \dots = \sum^{(q)}$$

を行う形をとる。

各変数とその標本の標準偏差で割る事によって、相関行列  $\sum_R^{(1)}$  に変換することは出来る。これにより各球体の大きさの差異をなくす事が出来る。p次元空間におけるq個の母集団から、q個の母集団の平均の差を考える変換により、N個のp次元観測値の間の最大変動に第一軸が存在し、ついで第二軸は2番目に大きな変動で、かつ第一軸に直交する形をとる。以下同様に軸を設定する。それぞれの係数は固有値に対する固有ベクトルであり、各母集団の重心を中心とする信頼限界の半径は、 $\chi^2$ 乗分布の限界点を変換したのもで近似出来る。

(注) 別名 多重判別関数(Multiple Discriminant Function, Tiederman & Bryan 1951)とか、多重判別分析(Multiple Discriminant Analysis, Coolley & Lohnes 1962)とか、正準判別分析(Canonical Discriminant Analysis 後藤 1970)等と呼ばれている。  
ここでは、Rao & Seal(1964)が命名した正準分析(Canonical Analysis)として紹介する。

#### < 解析目的 >

清浄度A系は何が原因として考えられるか、を解析してみたい。原因として考えられる要素をグループ毎に分類する。

- (1) レドール成分 (C、Si、Mn、P、S、Al) 群
- (2) 温度(出鋼)、時間(鋳込)、圧延比(スラブ厚/製品厚) 群
- (3) 合金系(HC Si-Mn、Fe-Si、Al)、終点C、Mn 群

これらの各群が上記データの分類により3分割(良、可、不可)したとき、清浄度の3分割に対応する説明変数の全グループの正準分析の距離とマハラリスの汎距離による計算方法で求め、そのグループ分けに及ぼす説明変数群の影響をみて、それぞれの影響度を検討する。

#### < 計算ステップ >

$W$  : 群内分散共分散行列、q個の母集団に対応する。

$$w_{ij} = \sum_{k=1}^q \sum_{\alpha=1}^{n_k} (x_{\alpha i}^{(k)} - \bar{x}_i^{(k)}) (x_{\alpha j}^{(k)} - \bar{x}_j^{(k)}) \quad (i, j = 1, 2, \dots, p)$$

$D_{kk'}$  : 第k群と第k'群の重心間のマハラリスの汎距離

$$D_{kk'} = \sum_{i=1}^p (x_i^{(k)} - \bar{x}_i^{(k)}) \cdot \sum_{j=1}^p w_{ij}^{-1} (x_j^{(k)} - \bar{x}_j^{(k)}) \quad (k > k') \quad (i, j = 1, 2, \dots, p)$$

$$\bar{x}_i = \frac{1}{n} \left( \sum_{k=1}^q n_k x_i^{(k)} \right) \quad (i = 1, 2, \dots, p)$$

$W$  の逆行列を  $W^{-1}$  とする。

$B$  : 群間分散共分散行列

$$b_{ij} = \sum_{k=1}^q n_k (x_i^{(k)} - \bar{x}_i) (x_j^{(k)} - \bar{x}_j) \quad (i, j = 1, 2, \dots, p)$$

行列  $W^{-1}B$  の固有値を求める。

$$|W^{-1}B - \lambda I| = 0, \quad I : \text{単位行列}$$

固有値は  $m$  個まで

$$\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_m > \lambda_{m+1} = \dots = \lambda_p = 0$$

この時の固有値に対応する固有ベクトルを求める。

$$W^{-1}Bx = \lambda x$$

$$Ax = \lambda x$$

とくと、次の標準固有値問題となる。

$$Ax = \lambda x$$

各固有ベクトルを

$$\mathbf{a}_1 = (a_{11}, a_{21}, \dots, a_{p1})', \mathbf{a}_2 = (a_{12}, a_{22}, \dots, a_{p2})', \dots, \mathbf{a}_p = (a_{1m}, a_{2m}, \dots, a_{pm})'$$

第  $j$  番目の判別関数

$$z_j = \sum_{i=1}^p a_{ij} (x_i - \bar{x}_i) \quad (j = 1, 2, \dots, m)$$

この判別関数が正準軸になる。

各観測値における判別得点 (正準平面における得点)

$$y_{\alpha j} = \sum_{i=1}^p a_{ij} (x_{\alpha i}^{(k)} - \bar{x}_i) \quad (\alpha = 1, 2, \dots, n_k, j = 1, 2, \dots, m, k = 1, 2, \dots, q)$$

95%信頼限界

Bartlett の  $\chi^2$  統計量

$$\chi_B^2 = \left\{ n - 1 - \frac{1}{2}(p+q) \right\} \cdot \log \left\{ \prod_{i=t+1}^m (1 + \lambda_i) \right\} \quad (t=0, 1, \dots, m-1)$$

自由度  $(p-t)(q-t-1)$

この検定を  $t$  について逐次反復する事により、何個の正準変数が有意であるかを

調べる。

第 k 群の重心に対して 95%信頼限界の円 ( 2 次元正準平面上 ) の半径

$$R^{(k)} = \sqrt{\frac{PC}{n_k}} \quad (k=1,2,\dots,q)$$

PC は自由度  $(p-t)(q-t-1)$  の  $\chi^2$  分布の 5% 点、 $n_k$  ( $k=1,2,\dots,q$ ) は各グループの サンプル数

$\chi^2$  分布の密度関数 ( n は自然数 )

$$f(x) = \frac{1}{2^{\frac{n}{2}} \Gamma\left(\frac{n}{2}\right)} x^{\frac{n}{2}-1} e^{-\frac{x}{2}} \quad (0 < x < \infty)$$

( x = 0 の時 f(x)=0 )

$\chi^2$  分布の限界点

d.f. \	10%	5%	1%
2	4.61	5.99	9.21
4	7.78	9.48	13.28
5	9.24	11.07	15.09
10	15.99	18.31	23.21
12	18.55	21.03	26.22

< 計算結果 >

固有値、ハートレット値、及びその有意性

各群	固有値番号	固有値	寄与率	累積寄与率	ハートレット値	自由度	有意性
レートル成分群	1	48.69	0.88	0.88	280.20	12	高度に有意
	2	3.32	0.06	0.94	96.63	5	"
温度時間圧延比群	1	1.08	0.68	0.68	53.30	10	高度に有意
	2	0.25	0.16	0.84	18.85	4	"
合金群	1	7.33	0.71	0.71	163.51	10	高度に有意
	2	1.45	0.14	0.85	63.87	4	"

各グループ毎の正準平面の重心点と信頼限界（半径）

分類	正準軸	レドール成分群		温度時間圧延比群		合金属群	
		重心	半径	重心	半径	重心	半径
グループ 清浄度良 Ga	$z_1$	0.82	0.377	- 0.21	0.377	- 0.40	0.377
	$z_2$	0.35		0.06		- 0.09	
グループ 清浄度可 Gb	$z_1$	- 0.001	0.387	0.39	0.387	0.05	0.387
	$z_2$	- 0.66		0.002		0.21	
グループ 清浄度不可 Gc	$z_1$	- 1.03	0.424	- 0.19	0.424	0.44	0.424
	$z_2$	0.35		- 0.08		- 0.13	

正準平面における各グループ毎のマハラリスの汎距離

分類		レドール成分群	温度時間圧延比群	合金属群
グループ A	グループ B	AB 間の距離	AB 間の距離	AB 間の距離
グループ Ga	グループ Gb	3.43	0.21	0.93
グループ Ga	グループ Gc	11.61	0.11	0.01
グループ Gb	グループ Gc	4.77	0.21	0.92

各正準軸

レドール成分群の正準軸

$$z_1 = 0.06781C - 0.01218Si - 0.00996Mn - 0.00807P - 0.31789S + 0.01676Al$$

$$z_2 = - 0.06213C - 0.07460Si - 0.01513Mn - 0.04573P + 0.10217S + 0.04848Al$$

温度時間圧延比群の正準軸

$x_1$  : 圧延比（鋼塊 / 製品） 、  $x_2$  : 出鋼温度 、  $x_3$  : 鍋中温度 、  
 $x_4$  : 鋳込時間 、  $x_5$  : 圧延比（スラブ / 製品）

$$z_1 = - 0.02007 x_1 + 0.09376 x_2 - 0.04282 x_3 + 0.09633 x_4 + 0.04522 x_5$$

$$z_2 = - 0.01080 x_1 - 0.00705 x_2 + 0.02638 x_3 - 0.00550 x_4 + 0.08069 x_5$$

合金属群の正準軸

$x_1$  : 終点 C 、  $x_2$  : 終点 Mn 、  $x_3$  : Si-Mn 、  $x_4$  : Al 、  $x_5$  : Fe-Si

$$z_1 = -0.17421x_1 - 0.01334x_2 + 0.00040x_3 - 0.00447x_4 + 0.001164x_5$$

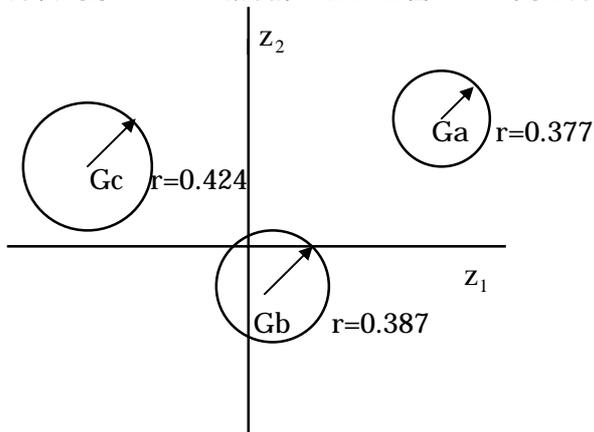
$$z_2 = 0.02455x_1 - 0.03996x_2 + 0.00021x_3 + 0.00354x_4 + 0.00084x_5$$

正準軸は正準分析による固有値、正準軸の係数は固有値に対応する固有ベクトルであり、各正準軸は互いに直交する。

(図8～図10)の各円は目的とした清浄度を3分類〔Ga(良)、Gb(可)、Gc(不可)〕した時、それに対応する説明変数群の各グループで、これは正準分析の評点となる集合体となる。円形は2次元正準平面( $z_1$ ,  $z_2$ 軸)に射影した各グループの95%の信頼限界の範囲を表す。

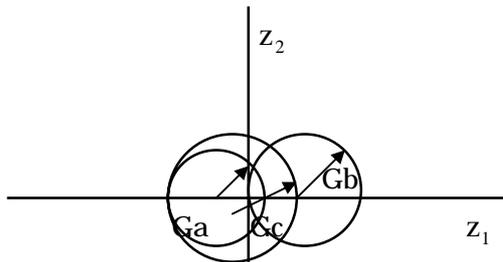
〔(図8～図10)では円の中心は各重心毎の位置、矢印は半径、(図8)のrは各グループの半径の値)〕

分析対象：レトル成分群における各グループ間の分布



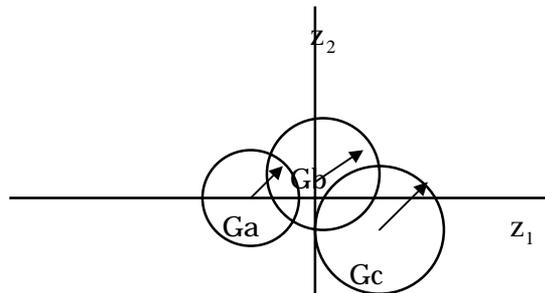
(図8)

分析対象：温度時間圧延比群における各グループ間の分布



(図9)

分析対象：合金群における各グループ間の分布



(図10)

<考察>

(図8)の場合、計算結果は重心間の距離においてレドール成分群がほぼ差が見られる形をしていた。88%の寄与率を占める $z_1$ 軸に対して  $G_a(良) > G_b(可) > G_c(不可)$ と成っている。出来れば $z_2$ 軸に対してもこれと同じ形(考え方によっては $G_a < G_b < G_c$ であっても良い)をしてほしいものであった。(図9)の温度時間圧延比群はこうした形を取っていない。すなわち温度時間圧延比群は清浄度A系のバラツキに寄与していないと判断できる。(図10)より合金群も少しではあるが、各グループ間に差が見られる傾向があった。レドール成分群では、第一正準軸 $z_1$ は88%の寄与率であり、これに対して第二正準軸 $z_2$ は6%の寄与率でありグループ間の変動は別な要因による変動が混じっているのではないかと考えても良い。

つまりこの場合第一正準軸だけを考えても良い。レドール成分群に対する第一正準軸は固有ベクトルを考えると最も大きいものはSであり、ついでCである。ここでSの係数はマイナスであることから清浄度A系を良くするためにはSを許容範囲迄下げて、Cは許容範囲内で上げる方が望ましいと解釈できる。合金群は第一正準軸 $z_1$ の寄与率が71%である所から、この第一正準軸で、Gaグループは $z_1$ 軸の低い位置にある所から、終点Cを上げるアクションをとることが望ましいと判断できる。

### 7 - 7 . 正準相関分析(Canonical Correlation Analysis)

変数群の群間の最大相関を求める事で、変数群同士の2つの群間の持つ意味合いを検討する事が出来る。各固有値に対する正準変量は互いに無相関である。正準変量を $\{u_k\}$ 、 $\{v_k\}$  ( $k=1,2,\dots,s$  ;  $k'=1,2,\dots,t$ )とおく時、2つの群はそれぞれの群内で無相関であり、お互いの間では、 $k=k'$ の一致する所だけ正準相関を持つ。この正準相関は固有値の数  $m = \min \{s,t\}$  だけ大きさの順に求められる。

<解析目的>

機械的性質が清浄度、及び経験上機械的性質に影響を及ぼすと考えられているチェック成分O2を含めた群との間の関係を調査する。

説明変数として清浄度、及び鋼質の性質からチェック成分O2を選ぶ。

$x_1$  : チェック成分O2、  $x_2$  : 清浄度A系、  $x_3$  : 清浄度B系、  
 $x_4$  : 清浄度C系

目的変数として機械的性質を選ぶ。

$y_1$  : Y.P.、  $y_2$  : T.S.、  $y_3$  : El

< 計算結果 >

固有値寄与率

固有値番号	固有値	寄与率	累積寄与率
1	5.02	0.58	0.58
2	2.34	0.27	0.85
3	1.29	0.15	1.00

第一番目の固有値に対する正準変量 (  $u_1$  ,  $v_1$  ) と正準相関係数  $R_1$

$$u_1 = -0.03312 x_1 - 0.02994 x_2 - 0.05030 x_3 + 0.04695 x_4$$

$$v_1 = -0.04949 y_1 + 0.05586 y_2 + 0.27461 y_3$$

$$R_1 = 0.62080$$

第二番目の固有値に対する正準変量 (  $u_2$  ,  $v_2$  ) と正準相関係数  $R_2$

$$u_2 = -0.20218 x_1 + 0.26313 x_2 + 0.74217 x_3 + 0.09587 x_4$$

$$v_2 = 3.71318 y_1 - 1.28694 y_2 + 0.28772 y_3$$

$$R_2 = 0.34631$$

第三番目の固有値に対する正準変量 (  $u_3$  ,  $v_3$  ) と正準相関係数  $R_3$

$$u_3 = -1.00900 x_1 - 0.03180 x_2 + 0.64337 x_3 - 0.11935 x_4$$

$$v_3 = 1.94392 y_1 - 3.05071 y_2 + 0.10930 y_3$$

$$R_3 = 0.10504$$

< 考察 >

寄与率が 58% の第一固有値に対する正準相関係数は 0.6208 である。清浄度及びフィッ成分 O<sub>2</sub> が良くなれば、概ね機械的性質も良く成って行く傾向が見受けられる。清浄度の中で機械的性質に影響を及ぼすものに A 系、B 系、C 系があるが、これらは余り大差がない。清浄度の群は機械的性質のうちどちらかと言えば、EI を大きく影響を及ぼしているものであることが推測できる。この EI が清浄度群より受ける影響は Y.P. T.S. のそれよりも約 5 倍位大きいことが推測できる。これは経験的にも他の分析からも妥当なものである事が分かる。

## 8. 多変量解析結果

### 8-1. 全体要約

鋼質の滑らかさや光沢を求めるにはSが影響し、必要であることが知られており、製品目的によっては鉄の美しさを求められる場合もある。一方Cは鋼質の硬さをもたらせる事が分かっており、硬い製品も必要である。機械的性質はいろいろな鋼質に関しても必要項目である。この機械的性質に影響しているものが清浄度であり、この内容を分析することによっていろいろなことが判明する。

清浄度をもたらす原因を幾つもの角度から分析した結果、おもにレドール成分系によるものであって、温度や時間、鋳型や板厚、圧延比などの影響はなかったものと解釈できる。清浄度A系と、B系は同じ要因ではなく、別な要因が絡んでいる事が判明した。

多変量解析を用いて生産管理への考え方を進めてきたが、これらの考え方のプロセスは、他の分野にも十分に適用可能である。

一方、金属内部におけることは、別な金属材料学によって理論的に、また実験を通して分かっていることもある。今回の分析はいくつかのものは、特に主成分分析から得られた機械的性質の最適値は、金属材料学から得られる実験値や理論推定値に、概算として合致している事が分かった。

また工場内の現実の生産過程のデータから新たな事柄も幾つか知り得る事が出来た。生産過程に注意すべきものとして、どのタイミングにどうすべきかは、別途生産指示書に記載したい。

### 8-2. 多変量解析として

一つ又は複数個の説明変数とそれに対応する目的変数として、重回帰分析は推定式あるいは決定式として用いる上で有効な手段であった。

他の一般報告書では、結果を少しでも良く見せる為、重相関係数は85%以上が多い。これは不適切なデータは削除したり、極めてフィットした良質だけのデータを用いたり、ダミー変数を加味したり、ある特定の期間のみを対象としたり、又は理論的に工夫して解釈を層別回帰とする場合がある。良く報告されている一般型重回帰分析で重相関係数を上げる為、変数を幾つも選んでいる場合も見あたるが、この時変数間の多重共線性が発生している場合もある様子で、十分な検討が必要と思われる。

現実の生産管理の生データでは、重相関係数はまれに90%以上のものもあるが、60%~70%程度が普通であってこれで良い方であり、結果はこの報告書に示した内容通りである。

主成分分析では、第一番目の固有値の寄与率が73%で、第二番目迄の累積寄与率が86%であり、主成分平面上の最大分散からの意味を考える上で、理想的な形をとる事が出来た。

因子分析ではこのデータでは固有値の寄与率がなだらかで、第二因子迄の累積寄与率が47%と鋼質に潜在する共通因子の複雑さが現れた形となった。

判別分析では判別可能な関数を作る為には、分離可能にする変数なら何でも良いと言う訳ではなく、他の解析において物理的意味を満足した有効な変数選択が必要と考えられる。

正準分析では、これまで他の事例が少なく比較対照に難しいものがあるが、今回のデータでは固有値等を考える限り、うまく解析が出来たものと判断出来る。理論では各グループの母平均が正準空間内で等しいと見なされるか否かの分析となるが、今回は目的変数を3グループに分類する事で、それに影響を及ぼす説明変数の立場と関連させて研究した。

正準相関分析も過去に他の事例が少なく、比較対照が難しいが、群毎に求められる相関と言う意味で第一正準相関（今回は固有値比率で58%）のみを取り上げれば、鋼質の全体性質を概略把握する事が出来、貴重な判断材料を提供してくれたものと考えられる。

いずれも多変量解析として、現実の生産管理データを解析し、意味を解釈する事が出来た。

### 8 - 3 . 解析結果の結論

この事例集では、今回のいろいろな解析を通してある一定の方向性を導き出す事が出来た。要点のみ簡素に表現したい。

（但し、成分値の範囲は、金属材料学より得られた表を用いる事が望ましい。）

- (1) 清浄度 A 系はレドール成分 S の影響が大きく、この S の値は製品として必要な最低限度の範囲としておく事が望ましい。
- (2) 清浄度 B 系はレドール成分 S の影響は認められず、合金 Fe-Si の影響が利いており、この値は許容範囲内で出来る限り上げることが望ましい。
- (3) 清浄度 C 系は、B 系の場合と同じく、合金 Fe-Si の影響があり、この値は上げる方が望ましい。
- (4) 金属の引っ張り試験の破面の状態を採点する判定では、合金 Fe-Mn と合金 Fe-Si の相互作用が強く影響している様である。
- (5) 機械的性質に影響を与えるとされているフィック成分 O<sub>2</sub> に関しては、鑄込み時間、圧延比の条件に左右され、かつ終点成分における C の値によって影響されており、この値は上げる方が望ましい。

### 8 - 4 . 解析結果を振り返る

生産計画の工程全体を眺め、変数相互間を調査するには、統計学の最も基礎で分かり易く表現し易いヒストグラムを始め、標準統計量による変数間の平均値の有意差検定、変数間の相関分析等は誰にでも分かり易く、これに生産工程の流れを理解すれば、良いと考えられる。個々の解析手法においては、理論関係を少し学んだ人が対象となる。分散分析や重回帰分析等は社会科学系や工学系出身者の人には理解が促進し易い。この重回帰分析も目的変数が複数個ある多変量重回帰分析や、相関の影響を少なくする Ridge 回帰分析なども注目されており、応用面で適用し易い。又旧来からの実験計画法等も魅力的である。

この事例集の主成分分析以降の解析には、数学特有の空間論の考え方が入ってくる為、事例研究される人には、その理解と適用に事前学習が必要となり得る。その他検討したものに共分散分析やクラスターやコホート分析などもあり、いずれも実用的と言える。

これらの考え及び理論は平行する他の学問と並び、場合によってはそれまで不明であった点が判明したり、困難と考えられて来たことが、計算上求められたりする事があり、優れた理論体系である事が証明できた。

#### 8 - 5 . 今後の課題

今回の事例集では、第2章における「解析の前提」で述べた間隔尺度や比尺度を中心として取り扱った。生産現場でのデータでは分類尺度や順序尺度のデータもあり得る。連続変数だけでは表すことが出来ない分野、とりわけカテゴリーで得られた変数には数量化理論や、ノンパラメトリック統計学、更に多次元尺度解析法等の分野があり、現実の生産工程への適用と言う面で研究すべき事は多く存在している。

#### 9 . 検討

この事例集で取り扱ってきた生産管理のデータに多変量解析を行うとき、計算上それらの変数間で何らかの形の分散共分散の逆行列を含んでおり、合わせてサンプリング上の変動に従う多くの要素を含んでいる。例えば重回帰分析に関して、一般型の場合、数式構成の理論上は、互いに相関を持たない各変数の要素から構成されていると仮定される。説明変数と目的変数の関係は重相関係数と呼ばれ、重回帰により説明された全分散は説明変数の各要素の各々独立した分散の成分に分割される、となる。このデータが独立、つまり無相関な場合とは、現実のデータでは存在しない。無相関なデータとは、基準化されたデータの一般的な線型変換、特に直交変換を必要とする。理論を構成する条件と得られたデータとは必ずしも合致しない。これらのことは最も古い文献で、Karl Pearson(1902)がこの問題を提起し、近年に至っては Lohnes(1966)が論じた文献がある。

こうした問題を研究して行く中で、Hotelling(1933)は得られた多数のデータから構成されるべき式を考える上で、データの最大分散に着目し、固有方程式を導き、これを解く上でラグランジュの乗数を使用して、最大化されるべき関数について、制約条件をいれ、係数で偏微分することにより固有値を求め、これから主成分の考えを提起した。この主成分分析の考えや適用例は幾つかの文献に見あたるが、概ねデータの分布と軸の意味的取り扱いが多い。

この事例集では、重回帰においては、上記問題点を少しでも回避する為、各変数の偏F検定に基づくステップワイズな段階式重回帰分析により、選ばれる変数を限定し少なくする事により、現実のデータと理論のギャップを埋め、主成分分析の考えと適用に至っては、主成分の得点がもとのデータに対してどの様にあるべきかを探求し、その妥当性に関して考察

した。

因子分析は Spearman(1902)が一つの一般因子を提唱したものに起源が見られ、その後 Thurstone(1947)が単純構造の原理を主張し、多因子モデルを提唱した。この因子分析は成分によって与えられる解を改良する考えに基づくが、そこで共通因子を導き出し、因子負荷量も求める。「この因子軸の解釈も軸を回転させる事により意味が理解し易くなる。」 Kaiser(1958) これが因子分析の主流となっている。最近では実に多くの枝葉となる因子分析の理論が出されている。軸の回転方法もいろいろ研究されており、いろいろなアプローチの仕方が可能となっている。医学生物学等に事例が見られ、特に心理学分野には事例が多い。この事例集では他の解析手法と比較検討しながら話を進めて来たので、理解しやすいものと思われる。

判別分析は母集団の判別の必要性から生じたもので、Pearson(1920)の研究から始まっている。多変量測定空間の異なった母集団から、各サンプルはその全体の測定空間から部分空間へ射影される。幾つかのサンプルはどの母集団から得られたものかを知りたい時、この分析手法は有益である。同じ時期 Mahalanobis(1922)はこの空間内に測度の概念を入れ、Mahalanobis の汎距離(Generalized Distance)として正当化させ後に普及した。幾つかの文献には応用事例も提案されている。今回はどのようなアクションをとれば良いのかを合わせて検討し提案してみた。

この判別分析の考えを拡張させて、複数個の母集団として、改良発展させたものが正準分析である。Bartlett(1947)は多重判別分析の考えを提唱した。複数個の変量に対して、任意の尺度を標本全体に通して変えても、結果に影響を及ぼすことがない、つまり空間内に存在する複数個の変量母集団を考え、これらのそれぞれの母集団を、ある平均値の回りに分散共分散行列により特徴付けられる球体( $n$ 次元)をして散布しているものと考え、各固有値に対する正準軸を求め、その意味合いを考察する所にある。

今回の鋼質の研究では変数間の意味合いも含めて、大変うまく適合させる事が出来た。

正準相関分析は Hotelling(1935)によって提唱され、Anderson(1960)により計算理論が構築された。この解析手法は、2個以上の変数から構成される変数群があり、その中で2つの変数群として定義された関数による群間の最大相関を求める所にある。固有値に対応した正準相関係数は、群間の情報を保持したもので、幾つかの変数群の合わせた傾向値を調べるのに都合が良い。この事例の鋼質の研究においては、正準相関係数が大きくはなかったが、概ね実用に耐えられると判断できる。

事例途中で適用して紹介している分散分析は、古くから良く知られているもので適用事例も多く出ており、一元配置のみ記載したが、傾向を調査する上で、二元配置も実施しており、 $F$ 的には多変量分散分析も有効性を調査する上でやっている。ただ多変量分散分析に至っては、数式モデル適用において、少し工夫を要する為、別途考慮したい。

## 10.まとめ

今回の事例集は、多変量解析の適用としての応用事例として生産管理の生データを取り上げ、幾つかの解析手法を適用し紹介した。頁数の制約もあり、他のいろいろな多変量解析手法や、考え方のプロセス全体（生産工程図、ヒストグラム、変数間の相関関係調査、及びそのプロット図他）を記載出来ず、又一部を除き計算手法も紹介出来なかったが、重要と思われるもののみ列挙した。

他の幾つもの研究報告内容を調査してみると、理論上の方法論をいろいろ論じているか、応用事例であっても、採集したデータに対して、一つ又は二つ程度の手法を当てはめ、その有効性を論じたものが多い。例えば因子分析一つを取り上げても細部の展開においては、いろいろな考え方が提案されており、その枝葉に対して因子分析としての理論だけか、又は数個の簡単な人為的データを分析し、紹介されている。

この事例集では、現在考えられている多変量解析の多くの手法の内、代表的と思われる一部の手法であり、合わせてその効用を紹介した。この応用事例の有効性は、生産管理への適用に関する上での考察で、鉄鋼生産過程の上でその品質を評価する時、工程から得られたデータをもとに、多変量解析でその効用を試みた。結果はこれ迄まで経験的に分かっていた事に関する裏付けや、新しい発見もあった。

鋼質の機械的性質における最適な生産環境や成分条件を求める事も出来、多変量解析の適用は大変有効な手段であると判断する事が出来る。

多変量解析の個々の解法の妥当性、有効性、新規性、それらの是非については、別途議論を有するものもある。多変量解析における考え方の新規性、計算理論、展開等を含めて個々には、別な論文にてまとめたい。

### 〔参考文献〕

- [1] 奥野忠一 他、多変量解析法、日科技連、(1971)
- [2] M.G.ケドール、浦昭二、竹並輝之共訳、多変量解析の基礎、サイエンス社、(1972)
- [3] H.L.シール、塩谷実訳、多変量解析入門、共立出版、(1970)
- [4] 河田敬義他、数理統計、裳華房、(1966)
- [5] 伊藤孝一、多変量解析の理論、培風館、(1969)
- [6] Lachenbruch,P.A. Discriminant Analysis when the initial samples are misclassified. Technometrics. vol.8.
- [7] Maxwell,A.E. Canonical variates analysis when the variable are dichotomous. Educ. Psych. Meas. vol.21, No2.(1976)
- [8] Cooley,W.W.and Lohnes, P.R., Multivariate Procedures for the Behavioral Sciences, Jon Wiley & Sons,(1962)

- [9] Cramer, H. *Mathematical Methods of Statistics*, Princeton University Press, (1961)
- [10] Draper, N.R. and Smith, H. *Applied Regression Analysis*, John Wiley & Sons, (1968)
- [11] Cochran, W.G. and Bliss, C.I. Discriminant functions with covariances, *Ann. Math. Statist.*, vol.19. (1948)
- [12] Bartlett, M.S. *Multivariate analysis*. *J. Roy. Statist. Soc.* (1947)
- [13] Bartlett, M.S. The goodness of fit of a single hypothetical discriminant function in the case of several groups, *Ann. Eugen.*, vol.16. (1951)
- [14] Claringbold, P.J. *Multivariate Quantal Analysis*. *J. Roy. Statist. Soc.* (1958)
- [15] Hotelling, H. A generalized T-test and measure of multivariate dispersion. *Proc. 2-nd Berkeley Symp. Math. Statist. and Prob.*, vol.23. (1951)
- [16] Cooley, W.W. and Lohnes, P.R. *Multivariate Data Analysis*, John Wiley & Sons. (1971)