現代人の統計 2

多変量解析法

柳井晴夫
高根芳雄 著

林知己夫 編

朝倉書店

多変量解析法

柳井晴夫
高根芳雄

現代人の統計 2

朝倉書店

著者略歴

柳井晴夫

1940年　東京に生れる
1965年　東京大学教育学部教育心理学科卒業
1970年　東京大学医学部保健学科疫学教室助手
1977年　千葉大学助教授（人文学部心理学教室）
教育学博士
現在に至る

高根芳雄

1945年　東京に生れる
1970年　東京大学文学部心理学科卒業
1973年　ノースカロライナ大学サーストン心理測定研究所留学
1977年　マッギル大学助教授（心理学部）
文学博士・Ph. D.
現在に至る

現代人の統計2

多変量解析法　　**定価　2000円**

1977年 9月30日　初版第1刷
1981年 1月20日　第6刷

著者　柳井晴夫（やないはるお）
高根芳雄（たかねよしお）

発行者　朝倉邦造

発行所　株式会社 朝倉書店
東京都新宿区 新小川町 2-10
郵便番号 162
電話 03 (260) 0141 (代)
振替口座 東京 6-8673 番

〈検印省略〉

政弘印刷・渡辺製本

3341-111802-0032

はじめに

心理学，社会学，政治学，医学などのいわゆる行動科学と呼ばれる分野では，多くの原因が相互に独立ではなく，たがいに作用しあいながらある事象に影響を与えていることが少なくない．例えば，ある大学の入学試験における，数学（x_1），物理（x_2），国語（x_3）の成績と入学後の成績（y）の相関が $r_{x_1y}=0.8$，$r_{x_2y}=0.4$，$r_{x_3y}=0$ であったとしよう．この場合，国語と入学後の成績の相関係数はゼロであるから，国語は入学後の成績を予測するのに全く有用とならないと考えるのは誤ちで，x_1，x_2，x_3 の間の内部相関係数の大きさによっては，x_1 と x_2 を用いるよりも x_2 と x_3 を用いる方が正確な予測ができることがある．

多変量解析とは，このように変数間の内部相関係数を分析して，ある事象の予測や判別を効率的に行ったり，変数間の背後に潜む因子をさぐる手法の総称で，近年，電子計算機の著しい発展に伴って，その重要性が広く認められてきたものである．

そして，その適用範囲も選挙予測，天候の予測，購売予測，適性診断，医学における計量診断，疾病の疫学的原因の探究など幅広く無数に存在する．

ここ数年，多変量解析の有用性の認識が深まるにつれ，和書，洋書を合わせると数十冊の多変量解析に関する書物が出版されているが，それらは，理論的に高度で理解するのにかなり骨が折れるもの，または，理論面について不十分な解説しかしていないものが多いが，本書は，その中で多くの例題を通して，ここ数年来の理論的発展をふまえながら多変量解析の理論をなるべく平易に取り扱うように試みたことに特色がある．そして，第2，4章においては，重回帰分析，判別分析における変数選択に関する新しい考え方，さらに第6，7章に

おいては，最近著しい理論的発展のみられる多次元尺度法と離散データの解析をとりあげたが，多次元尺度法を多変量解析の範疇に含めて，一つの書物で取り扱ったことは，これまでの類書にみられない斬新さがあるといえよう．

本書は，全部で七つの章と付録からなるが，第1章においては，多変量解析の基本概念と相関係数，平均，分散，相関係数のベクトルによる表現を与えて，第2章以降を読みこなすための必要最少限の基礎知識を与える．つづいて，第2章から第5章までは，重回帰分析，判別分析，主成分分析，因子分析という比較的オーソドックスな多変量解析の手法を多くの例題を通して説明し，第6章，第7章においては，第2章～第5章の知識を土台として，多次元尺度法と離散データを扱う技法の理論を展開した．本書を読むに必要とされる最少限の線型数学の予備知識と，第2章で割愛した正準相関分析の理論は付録で示した．さらに，多くの問題を与えることによって，本書の叙述で不十分なところを補うように配慮した．したがって，本書においては，クラスター分析を除く多変量解析の主要な手法はすべて網羅したつもりであるが，紙数の関係で，実際のデータを用いた応用例の分析は割愛してあるので，文献を通して，必要な応用例を補ってほしい．

本書の執筆は，第3,4,5章が柳井，第6章が高根，他の章と付録は，二人の共筆によるものであるが，全文を二人が眼を通し，意見を交換しあって全体の構成がすっきりするように十分配慮したつもりである．

なお，本書の全文を精読して有益な助言をして頂いた林知己夫(統計数理研究所所長)，松原　望(筑波大学助教授)，高木広文(東大医学部保健学科疫学教室)の三氏と，編集上でひとかたならぬお世話になった朝倉書店編集部に感謝の意を表したい．

1977年8月

柳　井　晴　夫
高　根　芳　雄

目　　次

1. 多変量解析の基本概念 ・・・・・・・・・・・・・・・・・・・・・・ 1

1.1 多変量解析とは ・・・・・・・・・・・・・・・・・・・・・・・・ 1

1.1.1 はじめに ・・・・・・・・・・・・・・・・・・・・・・・・・ 1

1.1.2 説明変数と基準変数 ・・・・・・・・・・・・・・・・・・・・ 2

1.1.3 変数を構成する4つの尺度 ・・・・・・・・・・・・・・・・・・ 4

1.1.4 データの種類 ・・・・・・・・・・・・・・・・・・・・・・・ 6

1.1.5 多変量解析法の分類 ・・・・・・・・・・・・・・・・・・・・ 7

1.2 相関と関連 ・・・・・・・・・・・・・・・・・・・・・・・・・・ 8

1.2.1 変数がいずれも間隔尺度の場合 ・・・・・・・・・・・・・・・ 8

1.2.2 一方の変数が間隔尺度，他方が名義尺度の場合 ・・・・・・・・ 15

1.2.3 変数がいずれも名義尺度の場合 ・・・・・・・・・・・・・・・ 16

1.2.4 変数がともに順序尺度の場合 ・・・・・・・・・・・・・・・・ 17

1.3 基本統計量のベクトルによる表現 ・・・・・・・・・・・・・・・ 18

1.3.1 分散と標準偏差 ・・・・・・・・・・・・・・・・・・・・・・ 19

1.3.2 共分散と相関係数 ・・・・・・・・・・・・・・・・・・・・・ 20

1.3.3 共分散行列と相関行列 ・・・・・・・・・・・・・・・・・・・ 21

1.3.4 合成得点の分散と標準偏差 ・・・・・・・・・・・・・・・・・ 22

1.3.5 合成得点の相関係数 ・・・・・・・・・・・・・・・・・・・・ 25

1.3.6 平均値と平均偏差得点 ・・・・・・・・・・・・・・・・・・・ 27

1.3.7 名義尺度変数の行列表示 ・・・・・・・・・・・・・・・・・・ 30

2. 重回帰分析法 ・・・・・・・・・・・・・・・・・・・・・・・・・ 33

2.1 重回帰分析の代数モデル ・・・・・・・・・・・・・・・・・・・ 33

2.1.1 説明変数が1つの場合 ・・・・・・・・・・・・・・・・・ 33
2.1.2 説明変数が2つの場合 ・・・・・・・・・・・・・・・・・ 35
2.1.3 説明変数が p 個ある場合 ・・・・・・・・・・・・・・・・ 36
2.2 重回帰分析の幾何学的モデル ・・・・・・・・・・・・・・・・ 37
2.2.1 説明変数が1つの場合 ・・・・・・・・・・・・・・・・・ 38
2.2.2 説明変数が2つの場合 ・・・・・・・・・・・・・・・・・ 39
2.2.3 説明変数が p 個ある場合 ・・・・・・・・・・・・・・・・ 40
2.2.4 基準変数が2つ以上ある場合 ・・・・・・・・・・・・・・・ 41
2.3 偏回帰係数と偏相関係数 ・・・・・・・・・・・・・・・・・・ 41
2.3.1 内部相関の変化と重相関係数 ・・・・・・・・・・・・・・・ 41
2.3.2 偏相関係数と部分相関係数 ・・・・・・・・・・・・・・・・ 44
2.3.3 偏回帰係数の幾何学的意味 ・・・・・・・・・・・・・・・・ 46
2.4 重回帰分析の確率モデル ・・・・・・・・・・・・・・・・・・ 48
2.4.1 説明変数が1つの場合 ・・・・・・・・・・・・・・・・・・ 49
2.4.2 説明変数が p 個ある場合 ・・・・・・・・・・・・・・・・ 51
2.4.3 説明変数の選択方法 ・・・・・・・・・・・・・・・・・・ 53
2.4.4 説明変数追加の打ち切り基準 ・・・・・・・・・・・・・・・ 55
2.4.5 適用例 ・・・・・・・・・・・・・・・・・・・・・・・・ 58
2.5 非線型回帰分析 ・・・・・・・・・・・・・・・・・・・・・・ 59

3. 判別分析・・・・・・・・・・・・・・・・・・・・・・・・・ 63
3.1 判別分析(二群の判別) ・・・・・・・・・・・・・・・・・・・ 63
3.1.1 説明変数が1つの場合 ・・・・・・・・・・・・・・・・・・ 64
3.1.2 説明変数が2つの場合 ・・・・・・・・・・・・・・・・・・ 65
3.1.3 重回帰分析との関連 ・・・・・・・・・・・・・・・・・・・ 67
3.1.4 説明変数が p 個の場合 ・・・・・・・・・・・・・・・・・ 69
3.1.5 尤度比とロジスティック曲線 ・・・・・・・・・・・・・・・ 71
3.1.6 母共分散行列が等しいと仮定しない場合 ・・・・・・・・・・ 73
3.2 重判別分析(多群の同時判別) ・・・・・・・・・・・・・・・・ 76

3.2.1 重判別分析の理論 ・・・・・・・・・・・・・・・・・・・・ 77
3.3 判別分析における変数選択 ・・・・・・・・・・・・・・・・・・ 79
3.3.1 説明変数選択の一般方式 ・・・・・・・・・・・・・・・・・ 80
3.3.2 適用例 ・・・・・・・・・・・・・・・・・・・・・・・・・ 83

4. 主成分分析法 ・・・・・・・・・・・・・・・・・・・・・・・・ 84
4.1 主成分分析の方法 ・・・・・・・・・・・・・・・・・・・・・・ 84
4.1.1 合成得点の分散最大による方法 ・・・・・・・・・・・・・・ 84
4.1.2 合成得点と各変数の相関を最大にする方法 ・・・・・・・・・ 87
4.2 変数の規準化 ・・・・・・・・・・・・・・・・・・・・・・・・ 89
4.2.1 共分散行列と相関行列 ・・・・・・・・・・・・・・・・・・ 89
4.2.2 規準化の一般的方法 ・・・・・・・・・・・・・・・・・・・ 91
4.2.3 主成分分析と非負定値行列 ・・・・・・・・・・・・・・・・ 93
4.3 外的基準がある場合の主成分分析 ・・・・・・・・・・・・・・ 94
4.3.1 外的基準が名義尺度の場合 ・・・・・・・・・・・・・・・・ 94
4.3.2 外的基準が間隔尺度の場合 ・・・・・・・・・・・・・・・・ 97
4.4 主成分分析から因子分析へ ・・・・・・・・・・・・・・・・・・ 99

5. 因子分析法 ・・・・・・・・・・・・・・・・・・・・・・・・・ 101
5.1 因子分析モデル ・・・・・・・・・・・・・・・・・・・・・・・ 101
5.1.1 共通因子数が1つの場合 ・・・・・・・・・・・・・・・・・・ 101
5.1.2 共通因子数が r 個の場合 ・・・・・・・・・・・・・・・・・ 104
5.1.3 因子負荷量と共通性 ・・・・・・・・・・・・・・・・・・・・ 104
5.1.4 因子分析モデルの行列による表現 ・・・・・・・・・・・・・ 105
5.1.5 共通性の定め方 ・・・・・・・・・・・・・・・・・・・・・ 107
5.1.6 共通因子数の定め方 ・・・・・・・・・・・・・・・・・・・ 108
5.1.7 反復法と Minres 法 ・・・・・・・・・・・・・・・・・・・ 110
5.2 共通因子軸の決定 ・・・・・・・・・・・・・・・・・・・・・・ 113
5.2.1 相関係数とベクトル空間 ・・・・・・・・・・・・・・・・・ 113

5.2.2 共通因子軸決定の一般方式・・・・・・・・・・・・・・・・ 115
5.2.3 各種の方法・・・・・・・・・・・・・・・・・・・・・・・ 116
5.2.4 最尤法と正準因子分析法・・・・・・・・・・・・・・・・・ 121
5.2.5 共通因子軸の回転・・・・・・・・・・・・・・・・・・・・ 123
5.3 因子得点の推定法・・・・・・・・・・・・・・・・・・・・・・ 125
5.4 因子分析の一般手順・・・・・・・・・・・・・・・・・・・・・ 127

6. 多次元尺度法・・・・・・・・・・・・・・・・・・・・・・・・ 129
6.1 多次元尺度法の概要・・・・・・・・・・・・・・・・・・・・・ 130
6.1.1 多次元尺度法とは・・・・・・・・・・・・・・・・・・・・ 130
6.1.2 方法の概要・・・・・・・・・・・・・・・・・・・・・・・ 131
6.1.3 適用の要件・・・・・・・・・・・・・・・・・・・・・・・ 135
6.2 最適化法・・・・・・・・・・・・・・・・・・・・・・・・・・ 138
6.2.1 データが比尺度で与えられている場合・・・・・・・・・・ 138
6.2.2 データが間隔尺度で与えられている場合・・・・・・・・・ 139
6.2.3 データが序数尺度で与えられている場合・・・・・・・・・ 142
6.3 個人差を考慮した多次元尺度法・・・・・・・・・・・・・・・ 150
6.3.1 方法の概要・・・・・・・・・・・・・・・・・・・・・・・ 150
6.3.2 誤差のない場合の解法・・・・・・・・・・・・・・・・・・ 152
6.3.3 誤差のある場合の解法・・・・・・・・・・・・・・・・・・ 152
6.4 その他の手法・・・・・・・・・・・・・・・・・・・・・・・・ 154

7. 離散データの多変量解析・・・・・・・・・・・・・・・・・・・ 156
7.1 外的基準のある場合・・・・・・・・・・・・・・・・・・・・・ 156
7.1.1 予測変数が名義尺度の場合の重回帰分析・・・・・・・・・ 156
7.1.2 予測変数が名義尺度の場合の重判別分析・・・・・・・・・ 160
7.1.3 基準変数が序数尺度の場合の数量化法・・・・・・・・・・ 162
7.1.4 あらゆる尺度水準の混在を許す一般化された正準相関分析 165
7.2 外的基準のない場合・・・・・・・・・・・・・・・・・・・・・ 168

7.2.1 クロス集計の数量化 ・・・・・・・・・・・・・・・・・・ 168
7.2.2 被験者と項目の同時数量化 ・・・・・・・・・・・・・・・ 174
7.2.3 あらゆる尺度水準の混在を許す一般化された主成分分析 ・・ 177

付　録 ・・・・・・・・・・・・・・・・・・・・・・・・・・・・ 181
問の解答 ・・・・・・・・・・・・・・・・・・・・・・・・・・・ 191
文　献 ・・・・・・・・・・・・・・・・・・・・・・・・・・・・ 200
索　引 ・・・・・・・・・・・・・・・・・・・・・・・・・・・・ 207

1 多変量解析の基本概念

1.1 多変量解析とは

1.1.1 はじめに

人間には生まれながらにして個人差があり，身長，体重から血圧，むし歯の数などの身体的側面，さらには，性格，興味，能力などの心理的側面に至るまで，個人差は多様である．

ここで，人間の頭のよさ(能力)について考えてみよう．「A君はB君より頭がよい」，「B君はC君より頭がよい」という二つの命題を是とした場合，いわゆる三段論法を適用すると，「A君はC君よりも頭がよい」ことになるが，A君とB君の頭のよさの比較が文科的能力，B君とC君の頭のよさの比較が理科的能力を重視して行われているとすれば，A君とC君の頭のよさの比較については，もはや何もいうことができない．

アメリカの心理学者サーストン(Thurstone, L., 1931)は，人間の能力が知能指数のように単一な指標で表わされるものではなく，異なった能力をいくつかの因子の組み合わせとしてとらえる能力の多因子説を唱えている．したがって職業の選択に際しては，自分の得意とする能力を生かせるような仕事を選ぶことが望ましいことになるが，能力だけが，職業選択の要因になることはきわめてまれである．

自動車セールスマンとしての適性を考える場合，大学時代にいくら優の数が多かったとしても機械への興味が弱くしかも性格的に共感性の低い人は，販売成績が低くなることが懸念される．

このように，自分の適性にあった職業を選択するには，自分の性格，興味，能力などを総合的に，しかも多元的に考慮することが望まれる．

ところで，医学の進歩とともに赤痢やコレラといったようなある病原体が体内に入らなければ絶対に発病しない，感染症と呼ばれる疾患は激減し，それにかわって，ガンや脳卒中などの非感染性の疾患が激増している．

このような疾患の原因は単一なものではなく，気候，地域，食習慣，体質など多くの要因が，病気の発生に複雑に関与してくる．

このように，ある事象(職業に対する適性，病気の発現)が，錯綜した複雑な要因(性格，興味，能力，体質，食習慣)によって規定されていると考えられる場合，それらの要因がどのように働き合って，事象に作用するかを見きわめることは容易なことではない．場合によっては，事象そのものが一元的でなく，多元的な側面を有していたり，概念的に規定し難いものであることが少なくない．

多変量解析とはこのような場合，事象そのものの多元的測定とその事象の背後にあると想定される要因の多元的測定から

(1)　事象を簡潔に記述し，

(2)　事象に対する要因の影響を査定し，

(3)　要因効果の結合法則を探りあてる

ための一連の統計的手法の総称で，電子計算機の発達とともに，ここ数年来，法律学，政治学，経済学，心理学，言語学，社会学，教育学，農学，生物学，人類学，医学などの分野で幅広く使用されるようになってきた．

1.1.2　説明変数と基準変数

ところで，事象および要因の多元的測定の結果得られるデータは多変量データと呼ばれ，事象に関するデータyを基準変数(または目的変数)，要因に関するデータxを説明変数として区別している．一般に同一の個体についての説明変数xと基準変数yのデータがともに揃っている場合，基準変数yは外的基準とも呼ばれ，この場合の典型的な分析方法はxの値によってyの値を推定(または予測)するものである．

なお，説明変数と基準変数という言葉は対で使用されるもので，経済学では外生変数と内生変数，数理統計学では独立変数と従属変数と呼ばれることが多い．

表 1 変数の相互関係

① 外的基準がある場合
説明変数⟶基準変数
(独立変数)⟶(従属変数) ⟶予測判別の問題
内的基準 外的基準

② 外的基準が潜在変数として与えられている場合
説明変数⟶基準変数(潜在変数)
③ 外的基準がない場合
説明変数(または基準変数) ⟶内的構造分析

ここで1つの個体について p 個の説明変数のデータ $x_1, x_2, \cdots, x_p$ が得られた場合，誤差成分 ε を考慮すると，それらの変動によって基準変数 y の変動がほぼ説明されると想定すれば，

$$y=\alpha_1 x_1+\alpha_2 x_2+\cdots+\alpha_p x_p+\varepsilon \qquad (1.1)$$

が成り立つ．上式の $\alpha_1, \alpha_2, , \cdots, \alpha_p$ は説明変数 $x_1, x_2, \cdots, x_p$ のそれぞれが y に与える影響の強さを示すパラメータで多くの個体について得られる $(y, x_1, x_2, \cdots, x_p)$ の組を分析することによってその推定値 $(\hat{\alpha}_1, \hat{\alpha}_2, \cdots, \hat{\alpha}_p)$ を求めることができる．こうして得られた推定値を重みとする重みづき合計点(合成得点 p. 24 を参照)

$$\hat{y}=\hat{\alpha}_1 x_1+\hat{\alpha}_2 x_2+\cdots+\hat{\alpha}_p x_p \qquad (1.2)$$

は基準変数の予測値となるもので，$\hat{y}$ が y の値に全体としてできるだけ一致するように $\hat{\alpha}_j$ の値を求めなければならない．

ところが実際に得られたデータはたまたま得られたものであって，基準変数と説明変数の関係の全貌を表わしているとは限らない．このとき誤差成分 ε に適当な確率分布(多くの場合，正規分布)を仮定すると，パラメータの推定値 $(\hat{\alpha}_1, \hat{\alpha}_2, \cdots, \hat{\alpha}_p)$ がどのような分布をするかを確率論的に導き出すことができる．この分布を手がかりに推定値がパラメータと平均的にどのくらい近いかを査定することが可能となる．

ところで (1.1) 式は，説明変数 x_j と基準変数 y の間に成立すると想定される関係をモデル化したもので線型モデル(linear model)と呼ばれるものである．

次に，事象または要因のいずれか一方に関してのデータが与えられている場合，これらの変数間の相互関連の強さ，つまり内的構造を分析する場合を考え

よう．これは外的基準のない場合の分析法に相当するもので，与えられた変数 $x_1, x_2, \cdots, x_p$ の相関関係より

$$x_j = a_{j1}f_1 + a_{j2}f_2 + \cdots + a_{jr}f_r + \varepsilon_j \qquad (j=1, \cdots, p) \qquad (1.3)$$

というモデルを想定するものである．上式の右辺の $(a_{j1}, a_{j2}, \cdots, a_{jr})$ $(f_1, f_2, \cdots, f_r)$ はともに未知のパラメータで，このようなモデルは双線型モデル(bilinear model)と呼ばれ，因子分析のモデルに相当する．しかし (1.3) 式の右辺の $f_1, f_2, \cdots, f_r$ は各個体について与えられるパラメータで，直接的には観測されない潜在変数(latent variables)であると考えられる．これを何らかの方法で推定したものを $\hat{f}_1, \hat{f}_2, \cdots, \hat{f}_r$ とすれば，これを用いて (1.3)式 を (1.2)式 に還元することができる．

誤差成分が正規分布するという仮定が常に多少「誤った」想定であるのと同様，線型モデル(あるいはいかなるモデル)も常に多少の誤りを伴った表現である．ただ多くの多変量解析の手法が線型モデルに基礎を置いているのはその取り扱いが容易なこと，すなわち線型数学の強力な理論がそのまま使えること，また多くの事象が近似的に線型モデルによってとらえられること，つまり線型モデルの強靱性(robustness)によるところが大きい．もちろん，第6章で述べる多次元尺度法の場合のように本質的に非線型モデルが採用される場合もあるし，§2.5 で触れる非線型回帰分析のような分析法も存在する．また，相関係数の項で述べるように，非線型的な関係を線型モデルによって近似することが事象の本質を見誤らせることもあるので十分注意することが必要である．

1.1.3 変数を構成する4つの尺度

同一のモデルであっても，調査や実験によって得られたデータの数としての性質によって適用する分析方法が異なってくるので，その性質を4種類に分けて説明しよう．

ある実体に数値表現を与えることを測定というが，測定は対象となる実体の1つのモデル化であって，数値表現が，実体に関する経験的関係をできるだけ正確に表現していなければならない．

身長や体重のようないわゆる物理的量の場合には，測定値 y^* (データ) とモデル値 y (定義量)の間には，ある程度の許容誤差の範囲内で

$$y=ay^*$$

という比例関係が成立する．このように例えば100 kgは50 kgの2倍というように数の比の相等性が意味を持ち，しかも絶対零点が存在する尺度を比尺度(ratio scale)という．これに対し，y^*とyの関係は線型でも，学力検査の成績のようにテストの結果(y^*)が0であっても，実際の能力(y)は0とは限らない場合もある．つまり，テストの得点の差には意味があるが絶対零点は存在しないわけで，この場合yとy^*の間には

$$y=ay^*+b$$

という関係が成立する．このような尺度を間隔尺度(interval scale)と呼ぶ．

次に心理学や社会学の調査において，しばしば用いられる評定尺度で‘好き’…3点，‘ふつう’…2点，‘嫌い’…1点を与える場合を考えよう．この場合3点と2点，2点と1点の間が等間隔であるという保証はない．このように順序のみが意味を持つ尺度が序数尺度(ordinal scale)と呼ばれるものである．

多くの変数について序数尺度で測定されたデータがある場合，例えばある変

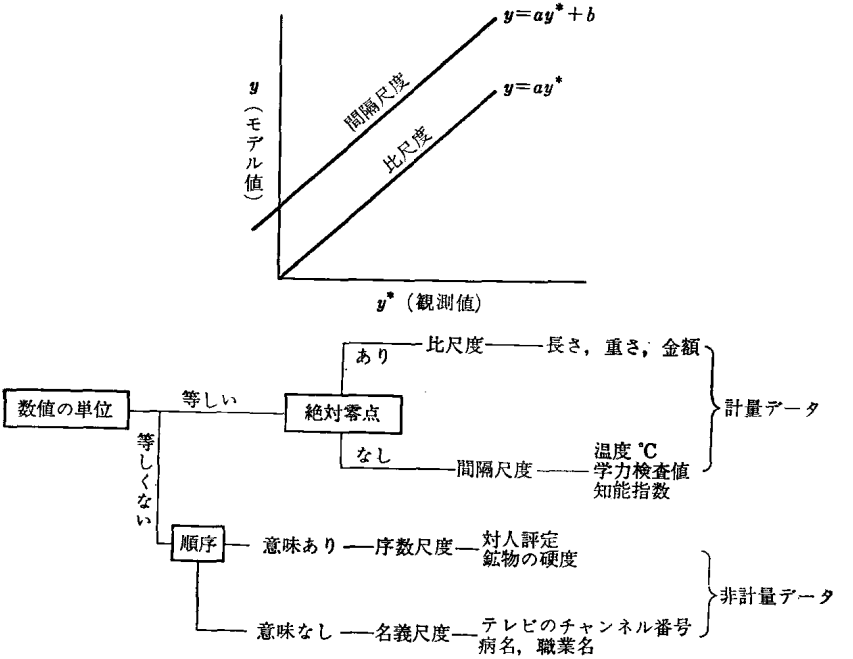

図 1 異なった尺度水準間の関連

数の 1, 2, 3, 4, 5 という順位に対して $y_1 \leq y_2 \leq \cdots \leq y_5$ という数値を対応させることができる．つまり序数データ $(y_i^* < y_j^*)$ に対して

$$y_i = f(y_i^*) \leq f(y_j^*) = y_j$$

となるような変換を単調変換 (monotonic transformation) と呼ぶ．

第4に，数値間の順序さえ意味を持たずテレビのチャンネル番号のように単なる記号としてしか意味を持たない尺度，すなわち，データ y^* とモデル y に1対1の対応関係のみが成立する尺度を名義尺度 (nominal scale) という．

なお，これらの4つの尺度の相互関連を示したのが図1である．

以上示した4つの尺度のうち比尺度と間隔尺度のデータを計量データ（メトリック・データ），序数尺度と名義尺度のデータを非計量データ（ノンメトリック・データ）と区別することがある．ただし，場合によっては評定尺度のような序数尺度のデータでも間隔尺度と同様に取り扱っても差しつかえないことがある．

1.1.4 データの種類

与えられたデータは，多くの場合，個体に関するさまざまな角度からの計測値，または評定値で，図2に示した矢田部・ギルフォード性格検査のようにプロフィール・データとして表わされる．

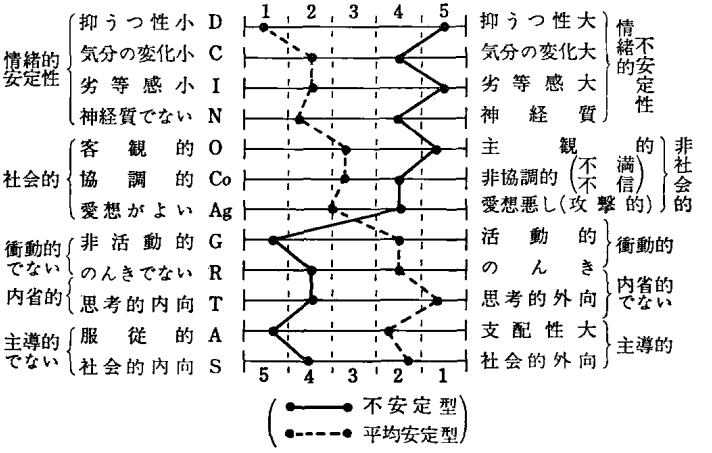

図 2　矢田部・ギルフォードの性格検査プロフィールの例

このような多変量データが多くの個体について得られた場合，これらの変数間の相関係数を求めることができる．ところで，相関係数は，変数間の類似性の程度を示す1つの指標であるが，多くの多変量解析の手法ではデータから導き出された相関係数，より根源的には相関係数が示すところの変数間の関連が解法の出発点となることが多い．これに対し，プロフィール・データによらず，直接，対象間の類似性の測度をデータとする分析法も存在する．

第6章で示す多次元尺度法は，さまざまな方法によって測定される対象間の類似性(あるいは非類似性)が分析の出発点となるものであるが，第2,3,4,5章で示す重回帰分析，主成分分析，因子分析，判別分析はプロフィール・データを分析の出発点としている．

1.1.5 多変量解析法の分類

これまで述べてきたことをまとめると，多変量解析の諸技法を分類する本質

表 2 多変量解析法の分類

		説明変数		基準変数	
		名義尺度	間隔尺度	名義尺度	間隔尺度
外的基準のある場合	重回帰分析	——	多数	——	1
	正準相関分析	——	多数	——	多数
	重判別分析（正準分析）	——	多数	多数	——
	(線型) 判別分析	——	多数	2	——
	数量化第一類	多数	——	——	1
	数量化第二類	多数	——	多数	——
外的基準のない場合	主成分分析	——	多数	——	(多数)
	変形主成分分析	——	多数	——	(多数)
	数量化第三類	多数	——	——	——
	クラスター分析	——	(多数)	——	多数
	因子分析	——	(多数)	——	多数
類似性指数を用いる	数量化四類，多次元尺度構成法，最小次元解析，潜在構造分析				

(1) (多数) は説明変数であっても，基準変数であってもよいことを示す．

(2) 因子分析はそのモデル構成において，一方は仮想的に与えられた潜在変数である．