二値データに基づく尺度の一次元性の評価の方法

塗　師　　斌*

Assessing the Unidimensionality of a Set of Dichotomous Variables

Akira NUSHI*

学力，知能，性格，適性，興味，態度など種々の能力や特性の測定やテストにおいて尺度は最も基本的に重要なものである。しかし心理測定における尺度として，物理測定における物差しや体重計のような客観的な尺度を構成することはなかなか難しい。従来，心理学的尺度が満たすべき基本的要件として主として取り上げられてきたのは妥当性と信頼性である。これに対してその尺度が測定している特性は確かに一つであるかという尺度の一次元性 (unidimensionality) については，特に問題にされない場合が多い。しかし近年，項目反応理論 (Item Response Theory，以下 IRT と呼ぶ) の適用が盛んになるにつれて，その基本的前提条件であるということもあって，尺度の一次元性が注目されるようになり，その評価の方法に関するさまざまな研究が行なわれるようになってきた。また Lumsden (1961), Hattie (1984, 1985), Hulin, Drasgow & Parsons (1983), Hambleton & Rovinelli (1986), Stout (1987) などにより次元性評価に関する研究の総覧もなされている。そこで本論文では，学力や知能や性格の測定データとして最も一般的な二値データをとり上げ，複数の二値データ (二値変数) によって構成される尺度の一次元性を評価するためのいくつかの主要な方法を，特に Hattie (1985) の review を参考にしながら説明して，今後の研究に資することを目的にする。

1.　一次元性を評価する必要性

Stout (1987) は，尺度 (あるいはテスト) の一次元性を評価することが何故重要であるかということに関して少なくとも次の三つの理由が考えられるとしている。

①たとえば数学的能力を測定しようとしているテストが言語的能力の影響を受けるといったように，ある能力の水準を測定しようとしているテストがそれ以外の能力の水準によって汚染されていたとすれば，妥当な測定とはいえない。

②一つの特性を測定しようとしているテストが実際には二つあるいはそれ以上の特性を測定していないかどうか。もしそうだとすれば得点が同じでも異なった意味をもつの

* 心理学教室 (Dept. of Psychology)

で個人差について言及できない。

③IRTに基づく標準的な諸方法を適用するためには，テストの一次元性が近似的にでも満たされなければならない。

また Hattie (1984) は「テストを構成する項目すべてがただ一つのものだけを共通に測定しているということは測定理論において最も critical で basic な仮定である」とし，個人の順序づけをしたり個人差について言及したり他の諸変数との関係を調べたりするためにはその項目群は一次元的でなければならないとしている。

以上からも明らかなように，テストあるいは尺度を構成する項目群が一次元的であるということは心理測定において何よりも基本的なことであり，得点の解釈や比較を行なうための不可欠の要件である。また Item bias, Equating, Adaptive testing 等の従来困難とされてきた問題に対して有力な方法を提供する IRT の適用を可能にするための基本的前提条件の吟味という意味でも，一次元性の評価は重要である。

2. 二値データの一次元性とは何か

学力テストにおいても知能テストにおいてもその基本データはそのテストを構成する各項目の正答か誤答かの二値データである。性格テストや各種の心理的特性の質問紙調査においても「はい」か「いいえ」かの二値データが基本データである場合が多い。またいうまでもなく IRT は基本的に二値データについての理論である。このように一般的でかつ重要な二値データの一次元性を評価するにはどのようにしたらよいのであろうか。そのためにはまず二値データの一次元性の定義を明らかにしなければならない。というのは，一次元性の評価の基準や方法や指標は何をもって一次元的とするかによって決まるからである。しかるに従来の一般的傾向として，Hattie (1985) も指摘するように，一次元性の定義を明確にしないまま，たとえば信頼性や内部一貫性 (Internal consistency) や等質性 (Homogeneity) の概念と混同して使われる傾向がある。そこでまずこの区別を明確にしておく必要がある。信頼性とは観測得点のなかで true score の占める比率を表すもので，この定義からして一次元性とは区別される。信頼性の測定のために再テスト法，平行テスト法，折半法，内部一貫性による方法などが用いられるが，この中で内部一貫性による方法が一次元性の概念と比較的近く，その一つである Cronbach (1951) の α 係数，あるいは二値データの α 係数とみなされる Kuder-Richardson の公式 20 (KR-20) が，従来，一次元性の指標として非常によく用いられてきた。その背景には Hattie (1985) も指摘するように，α 係数がランク(階数)が1の項目間相関行列と関連しているという考えがあった。しかし Green ら (1977) は一般因子が存在しなくても α 係数が高くなりうることを示した。また Green ら (1977) はシミュレーションによって，α 係数が項目数の増大や各項目に関連する因子数の増大に伴って大きくなることを示し，特に項目数の増大の影響を受けることが次元性の指標としての α 係数の主要な欠陥であると結論づけた。このように α 係数は一次元性の単調な関数ではないのである。等質性という用語は，Hattie (1985) によれば，一次元性の同義語として使われる場合と，項目間相関の類似性を表す場合の二通

りの意味で使われてきているとし，前者の使い方は redundant で confusing であり，後者の使い方は望ましくないとしている。これについては等者も同意見で，たとえば項目間の「等質性」が複数の次元でも生じる可能性が大きいことを考えれば，等質性という用語を使う場合にはどのような点で等質であるかを明確にすべきであると考える。以上述べてきたように，一次元性とは信頼性，内部一貫性，等質性とは区別して考えられなければならない。

一次元性とは，Lumsden (1961) によればすべての項目が同じ特性を測定していることであり，McDonald (1981, 1982) によれば局所独立 local independence の原理に基づくべきであり同じ能力の被験者群において項目間の共変動がゼロになることである。また Hattie (1985) によればデータの基底に一つの潜在特性 latent trait が存在しているものと定義される。これらの定義の要点は，すべての項目が同じ一つの特性をしかもそれだけを測定しているということである。しかし実際にはこのような定義からだけでは一次元性を評価するための具体的な基準や方法は一つに定まらず，一次元性のとらえ方の違いによっていくつかの方法が提案されている。そこで次節では，従来用いられてきた一次元性についての主要な考え方と評価の方法を取り上げて説明していく。

3. 二値データの一次元性を評価するための方法の概観

Hattie (1984, 1985) は一次元性を評価するために提案されてきた指標を総覧して 87 個取り上げ，それらを①回答パターンに基づく方法，②信頼性に基づく方法，③主成分分析に基づく方法，④線形因子分析および McDonald (1967) の非線形因子分析に基づく方法，⑤潜在特性モデル(近年これを IRT とよぶ場合が多い)に基づく方法の五つに分類している。①には Guttman (1944, 1950) の再現性係数，これを改良した Green (1956) の係数，Loevinger (1948) の等質性係数 index of homogeneity，②には α 係数，③と④の線形因子分析に基づく方法には第一固有値の大きさを利用する方法，⑤には 1 パラメタモデル (Rasch モデル) と 2 パラメタモデルおよび 3 パラメタモデルに基づく方法などが代表的なものとして取り上げられている。そして彼はこれらの指標についてモンテカルロ法のシミュレーションを用いて比較検討し，一貫して一次元と複数次元を区別できたのは，2 パラメタおよび 3 パラメタモデルを当てはめた後の残差（絶対値）の総和だけであったという結果を得ている。

Hambleton & Rovinelli (1986) は線形因子分析，非線形因子分析，残差分析 Residual analysis, Bejar (1980) の方法という四つの方法を取り上げてシミュレーション法によって比較検討を行ない，この中では非線形因子分析が最も promising な方法であるとしている。なおここでの残差分析とは Hattie の分類では⑤に相当するもので，テストデータに一次元の IRT モデルをあてはめてモデルパラメタの推定値を求め，理論的な予測値と観測値の間のズレの大小を分析する方法である。また Bejar の方法とは，Hattie の分類では⑤に含められるもので，テスト全体が測定している特性とは異なった特性を測定していると思われる一部の項目群を取り上げ，その項目群だけとテスト全体の項目群について

それぞれ3パラメタあるいは2パラメタモデルによる分析を行ない，取り上げた一部の項目群の困難度パラメタについて，この二種類の分析から得られる二通りの値の類似性を調べる方法である。

なお古くはなるが，Lumsden (1961) の総覧は後の心理測定学の発展を知るうえでも興味深いので紹介しておきたい。彼は一次元的なテストを作成するための方法を次の五つに分類している。それらは，①古典的な項目分析法 (item-test correlation の平均が最も高くなるような項目群を選ぶ方法)，②Loevinger の等質性係数，③相互依存性の基準 (interdependence criterion)，④回答パターン，⑤因子分析を用いる方法である。③の相互依存性の基準とは局所独立の原理に相当するものである。すなわち，一次元的なテストでは，true score が同じ被験者群において，ある項目に正答する確率は他のいかなる項目に正答する確率とも独立であるという原理に基づき，たとえば total score の同じいくつかの群を選んで，各群について項目間の独立性を χ^2 検定などを用いて調べていく方法である。また⑤では，項目間のテトラコリック相関行列のランクが1であること (unit rank)，第一因子の影響を除去したときの残差成分の大きさ，第一因子の分散の第二因子の分散に対する比率などが一次元性の基準として取り上げられている。そしてこれらの方法を比較吟味した結果，一次元性のための満足すべき基準は存在しないが，第一因子分散 (第一固有値) の第二因子分散 (第二固有値) に対する比率が (測定誤差は評価できないけれども) 多くの場合有効な指標になりうるとしている。

最近の傾向として，基本的には局所独立の原理に基づきそのための必要条件を設定して，項目間の分割表 contingency table がその条件を満たしているかどうかによって一次元性を調べる研究が盛んに行なわれている。その例として Holland (1981), van den Wollenberg (1982), Rosenbaum (1984), Holland & Rosenbaum (1986), Stout (1987) らの研究があげられる。たとえば Stout (1987) の方法は，あるテストを少数の M 項目からなる下位テスト (一次元性を評価するために用いるという意味で assessment subtest と呼ぶ) とその他の n 項目からなる下位テスト (被験者を得点別にグルーピングするために用いるという意味で，partitioning subtest と呼ぶ) に分けたとき，もしそのテストが一次元的であれば，partitioning subtest で同じ得点 k の集団では assessment subtest のどの二項目の共分散の値もゼロに近いという，式 (1) で示される基準を用いている。

$$\frac{1}{M(M-1)}\sum_{n+1\leq i\neq j\leq n+M}\mathrm{Cov}\,[U_i,\ U_j|\sum U_m=k]\to 0 \qquad (1)$$

以上，一次元性を評価するための各種の方法を概観してきたが，これらの方法のなかでこれまでまた現在でも最も多く使用されているのは，線形因子分析に基づく方法である。これと基本的には同じ方法である主成分分析がその代わりに使われる場合もある。非線形因子分析は Hattie (1985) も指摘するように有望な方法ではあるがパラメタ推定のための計算法が非常に難しいということもあって，実際にはほとんど使われていない。α 係数は現在でも一次元性と混同して使われることがあるが，前述したような理由でこの二つの概念は異なるので，はっきり区別して使用すべきであろう。Loevinger の等質性係数も含め

て回答パターンによる方法は直接的でわかりやすい方法であるが，系統誤差と偶然誤差の区別ができないことや，複数の次元でも同じ重みでテスト得点が合成される場合には一次元とみなされてしまうことや，大量のデータの処理にはあまり向いていないなどといった欠点のために，一次元性の評価の方法としてはあまり使われていない。IRT に基づく方法は，先にも述べたように，局所独立の原理に基づく方法も含めて研究が盛んであり，将来最も有望な方法であると思われるが，現段階では実際場面であまり使われていない。本論文ではこれらの方法の中から，最も多く使用されている線形因子分析（以下，単に因子分析と呼ぶ）に基づく方法，最も直接的でわかりやすい回答パターンによる方法そして将来有望な IRT に基づく方法を取り上げて，次節以降で各種の具体的な方法を解説していく。

4.　因子分析に基づく方法

この方法は基本的には項目間の相関行列あるいは分散共分散行列の最大分散を説明する第一因子の分散すなわち第一固有値の大きさがテスト得点の全分散に占める比率に基づく方法である。一般的な手続きとしては，まず項目間の ϕ 係数あるいはテトラコリック相関係数の相関行列を求め，その対角成分に共通性の推定値として，各行成分の相関係数の絶対値の最大値あるいは SMC(重相関係数の平方)を代入し，その行列について因子を抽出して，各因子特に第一因子の固有値を調べ，非常に大きい場合一因子性が高いと判定するのである。ここで問題になるのはその判定基準である。これについては，いずれも特に明確な理論的な根拠がないままつぎに述べるようないくつかの基準が提案されている。まず Lumsden (1961) は先にも述べたように第一固有値と第二固有値の比率が，どの程度測定誤差の影響を受けるかわからないにしても最も有効であるとしている。また Lord (1980) は判定のための rough な基準として，Lumsden と同様第一固有値と第二固有値の比率を提案し，それと同時にさらに第二固有値が第三固有値以降よりもそれほど大きくならないことを付け加えている。これを指標化すれば，第一固有値と第二固有値の差を第二固有値と第三固有値の差で割ることになるであろうと Hattie (1985) は述べている。Hambleton & Rovinelli (1986) は判定のための一般的な手続きとして，固有値の大きさを順にプロットしたときに急に落ち込む前までの固有値の数を有意味な因子数にする方法（“elbow” in the plot による方法）を挙げている。しかしこの方法は elbow が見られないときには適用できない。また第一固有値のテスト得点の全分散に占める割合がどのくらい大きければ一次元的といえるのかについても特に明確な基準はなく，たとえば少なくとも 40% は必要であるという意見や，あるいは Reckase (1979) のように一次元の IRT モデルのパラメタ推定値が実際に第一因子をどの程度反映できるかという点では少なくとも 20% は必要であるとする意見などがある。

以上述べてきた方法のいずれも固有値の大きさがどのくらい大きければ一次元的であるといえるのかについての主観的な判断が要求されるが，これを客観的に行なう方法もある。実際によく用いられている方法として，相関行列の対角成分に 1 を代入したときすなわち主成分分析を行なったときの 1 よりも大きな固有値の数を因子数とする方法 (Kaiser,

1970) がある。また Linn (1968), Humphreys & Ilgen (1969), Humphreys & Montanelli (1975) らによる parallel analysis と呼ばれる方法がある。この方法は実際のデータと項目数と被験者数が等しいランダムな標準正規乱数 z からなるデータ行列を作成し，この各成分を $z=0$ を基準にしてそれより大きい場合を 1，小さい場合を 0 というように二値化し，これによって作成されるデータ行列の項目間の積率相関（ϕ 係数）の相関行列を因子分析（共通性の推定値として SMC を使用）することによって得られる第一固有値の大きさを，実際のデータの因子分析結果の因子数を決める際の cutoff score として用いる方法である。たとえば実際のデータから得られた第二固有値がランダムなデータ行列から得られた第一固有値よりも小さくなったとすればその実際のデータは一次元的であるということになる。すなわちランダムなデータからでも得られる程度の大きさの固有値は有意味ではないとみなすわけである。Drasgow & Lissak (1983) によれば，parallel analysis は適切な因子数を決定するうえで非常に有効であることが確認されているとされる。この方法を多少変形したものとして modified parallel analysis（以下，MPA と略す）と呼ばれる方法が，Drasgow & Lissak (1983) によって提案されている。この方法は，実際に得られた二値のデータから項目間のテトラコリック相関行列を求め，これを因子分析（共通性の推定値として相関係数の最大値を使用）することによって得られる固有値と，一次元性を満たすように作成した実際のデータと同じ大きさの人工データについて，同様な因子分析を行なうことによって得られる固有値の大きさを比較して，もし実際のデータの第二固有値が人工データの第二固有値とほぼ等しければ一次元性が高いと判定する。なお一次元性を満たす人工データを作成する方法としては，実際のデータについて適切な一次元の IRT モデルを適用してそのモデルの項目パラメタを LOGIST などを用いて推定し，その推定値を適用した項目特性曲線を用いて，発生させた標準正規乱数（被験者パラメタ θ に相当する）のそれぞれについて各項目の正答確率を求め，このそれぞれについて発生させた 0 から 1 までの一様乱数がそれに対応する正答確率より低ければ 1（正答），高ければ 0（誤答）を割り当てるという方法が使われる。このように MPA は parallel analysis とは逆に一次元性の高い人工データをモンテカルロ法によって作成して比較を行なうわけであるが，固有値の大きさの比較の際どの程度等しければ一次元的といえるのかという点で，主観的な判断を要求される場合もある。

二値データについて，あるモデルのもとでの推定量が χ^2 分布することを利用して，一次元的モデルと実際のデータとの適合度 goodness of fit を検定する方法として，最尤解を用いる Bock & Lieberman (1970) の方法，一般化された最小二乗解 generalized least squares を用いる Christoffersson (1975) や Muthen (1978) の方法がある。しかしこれらの方法はいずれも計算が非常に heavy で，実用的に計算可能な最大の変数の数は Bock & Lieberman の場合 10〜12，Christoffersson と Muthen の場合 20〜25 に限定される。また Jöreskog (1978) の方法は，二因子を仮定した場合の χ^2 値と一因子の場合の χ^2 値の差が χ^2 分布することを利用して，二因子を仮定した場合にどの程度適合度がよくなるかをみることによって一次元性を調べる。この方法に関して Hattie (1985) は，この χ^2 値は標本が多くなるとすぐ有意になりやすいので，trivial な差でも仮説の棄却に導きやす

いことを指摘している。

McDonald (1982) は，一因子がデータに適合するかという仮説の検定は標本の数さえ多ければ χ^2 値が有意になる傾向があることを指摘し，第一因子を除去した残差共分散の大きさのほうが，データに対する一因子モデルの misfit の程度を判断する基準として適当であるとしている。

以上，因子分析に基づく各種の方法を述べてきたが，その他にも，Tucker & Lewis (1973) の一因子に関連した分散とテスト全体の分散の比率に基づく方法，McDonald (1980) の theta という値を一次元性の指標として用いる方法，Green et al. (1977) の共通性に基づく方法などがある。Green et al. (1977) の方法は，項目間の共通因子が一つであるとすれば因子負荷はそれぞれの共通性 h_i^2, h_j^2 の平方根に等しく，また項目間の相関係数 γ_{ij} はそれぞれの因子負荷の積すなわち共通性の平方根の積に等しくなることを利用し，式 (2)

$$u=\sum\sum|\gamma_{ij}|/\sum_{i\neq j}\sum(h_i^2 h_j^2)^{1/2} \qquad (2)$$

で表される u を一次元性の指標とするものである。この指標の大きな問題点は，Hattie (1985) も指摘するように，次元性についての正しい知識が前提となる共通性がわかっていなければならないことである。

5.　回答パターンに基づく方法

Lumsden (1961) は表1のようにテスト得点からその人の回答パターンが明らかになるような回答パターンを一次元的であると定義している。

表1　5項目からなる一次元的なテストの回答パターン

テスト得点	項目				
	1	2	3	4	5
0	*F*	*F*	*F*	*F*	*F*
1	*P*	*F*	*F*	*F*	*F*
2	*P*	*P*	*F*	*F*	*F*
3	*P*	*P*	*P*	*F*	*F*
4	*P*	*P*	*P*	*P*	*F*
5	*P*	*P*	*P*	*P*	*P*

(P=Pass,　F=Fail)

これは Guttman (1944, 1950) の絶対尺度，perfect scale,完全ガットマンスケールあるいはスケイラブルな尺度に相当するもので，Guttman の再現性係数（表1のような規則的な回答パターンからズレた回答数を全回答数で除した値を1から減じた値）が1になるような回答パターンである。Guttman (1945) は再現性係数が0.90以上である場合一次元的とみなせるとしている。しかしこの再現性係数は項目困難度の影響を受けたり，cutting point を決めるのが面倒であるなどといった欠点があるので，これを克服するものとして Green (1956) の指標が提案されている。これは二値データ項目について，困難度や通過率の高い（あるいは低い）順に項目を並べ換え，隣り合った項目間のみについて分割表を求め，たとえば難しいほうの項目は正答して易しいほうの項目を誤答した場合を規則的なパターンからズレたエラーとみなし，全分割表でのエラーの総数を全回答数で除した値を1から減じた値を一次

元性の指標とするものである。

Loevinger (1948) の等質性係数も表1のような絶対尺度の概念に基づく指標である。これはテストの項目間の共分散の総和を，項目間で可能な最大の共分散の総和で除した値に等しいことが示される（池田，1973)。すなわち，表1のようなパターンが成り立つときには，難しいほうの項目 j（正答率を P_j とする）に正答して易しいほうの項目 i（正答率を P_i とする）に誤答する人はおらず，また難しいほうの項目に正答した人は必ず易しいほうの項目にも正答するはずであるから，項目 i と項目 j の同時正答率を P_{ij} とすれば $P_{ij}=P_j$ が成り立ち，分割表は表2のようになる。いまそのテストの n 項目を易しいものから $P_1 \geqq P_2 \geqq \cdots \geqq P_i \geqq \cdots \geqq P_j \geqq \cdots \geqq P_n$ という順に並べ換えて新しい添字を用いると，表1のパターンが成り立つときすなわちテスト得点の分散 σ^2 が最大になるときの値 $\sigma^2_{\max}$ は，テスト得点の平均を μ とすれば，式 (3) で表すことができる。したがって項目間で可能な最大の共分散の総和は式 (4) で表される。

表2　絶対尺度が成り立つときの分割表
($P_i > P_j$ の場合)

		項目 j 正答	項目 j 誤答	
項目 i	正答	P_{ij}	P_i-P_j	P_i
項目 i	誤答	0	$1-P_i$	$1-P_i$
		P_j	$1-P_j$	

$$\sigma^2_{\max}=2\sum_{i=1}^{n} iP_i-\mu(1+\mu) \tag{3}$$

$$\text{最大共分散}=\sigma^2_{\max}-\sum_{j=1}^{n} P_j(1-P_j) \tag{4}$$

これに対して実際に得られるテストの項目間の共分散の総和は式 (5) で表される。この式 (5) を式 (4) で除した式 (6) が Loevinger (1948) の等質性係数なのである。

$$\text{実測される共分散}=\sigma^2-\sum_{j=1}^{n} P_j(1-P_j) \tag{5}$$

$$\text{Loevinger の等質性係数}=\frac{\sigma^2-\sum_{j=1}^{n} P_j(1-P_j)}{2\sum_{i=1}^{n} iP_i-\mu(1+\mu)-\sum_{j=1}^{n} P_j(1-P_j)} \tag{6}$$

以上述べてきたような回答パターンに基づく方法は，近年盛んになってきた order analysis という方法 (Bart & Krus, 1973; Krus & Bart, 1974; Cliff, 1975 など）のなかでも取り上げられている。たとえば Loevinger の等質性係数は，Cliff (1977) の C_{t3} に相当する。order analysis とは二値データについて項目間の支配関係 dominance relations を利用して，項目間や被験者間でガットマンスケールのような順序関係あるいは階層関係を求める方法である。ここで支配関係とは，たとえば項目 i では正答するが項目 j では誤答するという関係が強いと，項目 j は項目 i を支配しているという意味で使われる。表2のような場合には項目 j は項目 i を完全に支配しているということになる。Wise (1983)

は二値データの次元性の評価に関して，order analysis と因子分析の比較を行なっている。彼は一因子の場合と二因子の場合についてそれぞれ項目困難度の散らばりを大，中，小と変化させた人工データのそれぞれの場合について，Krus & Bart (1974) の方法，および Cliff (1977) の C_{t1} と C_{t3} を用いた Reynolds (1981) の方法を適用した結果，Krus & Bart の方法と C_{t1} を用いた方法はデータの因子構造をほとんど反映できなかったが，C_{t3} を用いた方法はすべてのデータセットの因子構造をかなり正確に反映したという結果を得ている。しかし C_{t3} も因子間で相関がある場合には異なった因子の因子負荷を識別できないことがあるとされる。なお Krus & Bart の方法と C_{t1} を用いた方法が因子構造を反映できなかった理由として，項目間で困難度が類似している場合には二つの項目が同一次元であったとしてもその支配関係を表現できず，逆に困難度が非常に異なる場合には異なる次元であるにもかかわらず支配関係を生じやすい傾向があるからだとしている。これは order analysis における基本的な問題でもあり，Wise や Krus & Krus (1980) が指摘するように二つの項目を同時に正答あるいは誤答するという親近関係 proximity relations を分析のなかにうまく取り込むことが今後の課類といえよう。実際に，Wise & Tatsuoka (1986) は親近関係をも分析対象とすることによって，order analysis と因子分析の結果が非常に一致することを示している。

6. 項目反応理論（IRT あるいは潜在特性理論）に基づく方法

二値データ項目の場合，項目得点の潜在特性 θ への回帰曲線でありかつ項目 i に対して正答する確率を表す項目特性曲線 (item characteristic curve) $P_i(\theta)$ は，潜在特性が一次元の場合で三種の項目パラメタを仮定する場合には，式 (7) あるいは式 (8) で表される。式 (7) は3パラメタロジスティックモデル，式 (8) は3パラメタ正規累積（あるいはオージブ）モデルと呼ばれ，式中の a_i, b_i, c_i はそれぞれ項目 i の識別力，困難度，偶然正答 (guessing) の確率を表す項目パラメタである。また式 (7) の D は定数でこの値が 1.7 のとき，式 (7) と式 (8) の値はほとんど同じで，θ の全域にわたって 0.01 以上の差は生じないとされる。

$$P_i(\theta)=C_i+(1-C_i)[1+e^{-Da_i(\theta-b_i)}]^{-1} \quad \text{（ロジスティック曲線）} \qquad (7)$$

$$P_i(\theta)=C_i+(1-C_i)\int_{-\infty}^{a_i(\theta-b_i)} \frac{1}{\sqrt{2\pi}} e^{-z^2/2}dz \quad \text{（オージブ曲線）} \qquad (8)$$

三種の項目パラメタの中の c (guessing パラメタ) を0とみなすモデルは2パラメタモデルと呼ばれ，その項目特性曲線は式 (7)，式 (8) から $C_i+(1-C_i)$ の部分だけを除いた式で表現される。また c を0とみなすだけでなく，識別力パラメタ a の値が1でどの項目も同じであると考えるモデルは1パラメタモデルあるいは発案者の名前にちなんで Rasch モデルと呼ばれ，その項目特性曲線は式 (7)，式 (8) から $C_i+(1-C_i)$ の部分と a を取り除いた式で表現される。このような一次元の IRT モデルにおける最も基本的な仮定は，

局所独立の仮定である。局所独立とは Lord & Novick (1968) によれば「その潜在特性だけが唯一の重要な因子であって，被験者のその値がわかれば被験者の挙動が統計的独立という意味でランダムになってしまう」という状態で，具体的には潜在特性を一定の値に固定したときに任意の項目 i への回答と項目 j への回答が統計的に独立になるという意味である。

以上のような一次元の IRT モデルに基づきテストデータの一次元性を評価するための基本的な方法は，データにある適切な IRT モデルを当てはめてそのモデルの項目パラメタおよび被験者パラメタ θ の推定値を求め，それをモデル式に代入して得られる理論的な正答確率と観測値とのズレあるいは適合の度合いを summarize することである。たとえば推定された θ の値に基づいて被験者をいくつかの群に分類し，各群について各項目の実際に得られた正答率と，用いた IRT モデルから予測される正答率とを比較するという方法がよく使われる。Wright & Panchapakesen (1969) は，項目のセットがモデルに適合するならばそれは一次元性を示す証拠であるとしている。また局所独立の原理を応用した例として，Lord (1953) はある得点水準の被験者をすべて選び出して，いかなる二項目に対する回答も独立であるかどうかを決定するために χ^2 値を求め，その和がやはり χ^2 分布することを利用して局所独立の仮定を調べる方法を提案している。

7. まとめ

以上述べてきたように，尺度あるいはテストの一次元性の評価に関わる方法は実にたくさんの種類があることがわかる。しかしその中でどの方法が最もすぐれているかについては，Stout (1987) も指摘しているように心理測定学者の間でも一致した見解がなく，実際多くの人は満足すべき方法がないと感じているのである。この最大の理由は，これまで説明した各種の方法がそれぞれの立場から指標を定義していたことからも明らかなように，一次元性をとらえる多くの観点があり，それだけ一次元性という概念が本来的に明確でないことによるのであろう。それにもかかわらず一次元性の評価は，一方では IRT の基本的仮定の吟味のために，また他方では尺度が備えるべき基本的要件の吟味という意味で，その必要性が非常に高いのである。これを解決する一つの方法は，それぞれが一次元性を評価する目的を明確にした上で，そこで要求される種類の一次元性にかなった指標や方法を用いることであろう。たとえば IRT の基本的仮定の吟味が目的であれば，データがどの程度一次元的であれば（あるいは多次元的でなければ）モデルパラメタの正しい推定が可能かという推定の robustness の観点から方法を選ぶことになろう。また項目間の一次元的な階層構造を見い出すことが目的であれば order analysis，一次元的尺度構成のための各項目の重みの算出が目的であれば因子分析的方法が適切であろう。その一方で，物理測定とは違ってここで問題にしている心理測定の場合に，極端に一次元的な尺度は果たして妥当な尺度といえるのかという本質的な問題もある。たとえば同じような質問を何回も繰り返せば一次元性はたしかに高くなるのである。しかしこのような尺度が妥当な尺度でないことはいうまでもないであろう。心理測定の尺度の場合一般に，それを構成する項目

は項目間で共通な特性と同時にその項目独自のものも測定しており，ある程度幅広さをもっているのが特徴である。このようにそもそもある程度は多次元的であるべき尺度に対して一次元性を判定することには，本来的な困難さが含まれていることが理解されよう。

参考文献

1) Bart, W. M. & Krus, D. J., 1973, An ordering-theoretic method to determine hierarchies among items. Educational and Psychological Measurement, 33, 291–300.
2) Bejar, I. I., 1980, A procedure for investigating the unidimensionality of achievement tests based on item parameter estimates. Journal of Educational Measurement, 17, 283–296.
3) Bock, R. D. & Lieberman, M., 1970, Fitting a response model for n dichotomously scored items. Psychometrika, 35, 179–197.
4) Christoffersson, A., 1975, Factor analysis of dichotomous variables. Psychometrika, 40, 5–32.
5) Cliff, N., 1975, Complete orders from incomplete data: Interactive ordering and tailored testing. Psychological Bulletin, 82, 289–302.
6) Cliff, N., 1977, A Theory of consistency of ordering generalizable to tailored testing. Psychometrika, 42, 375–399.
7) Cronbach, L. J., 1951, Coefficient alpha and the internal structure of tests. Psychometrika, 16, 297–334.
8) Drasgow, F. & Lissak, R. I., 1983, Modified parallel analysis: A procedure for examining the latent dimensionality of dichotomously scored item responses. Journal of Applied Psychology, 68, 363–373.
9) Green, B. F., 1956, A method of scalogram analysis using summary statistics. Psychometrika, 21, 79–88.
10) Green, S. B., Lissitz, R.W., & Mulaik, S. A., 1977, Limitations of coefficient alpha as an index of test unidimensionality. Educational and Psychological Measurement, 37, 827–838.
11) Guttman, L., 1944, A basis for scaling qualitative data. American Sociological Review, 80, 139–150.
12) Guttman, L., 1945, A basis for analizing test-retest reliability. Psychometrika, 10, 255–282.
13) Guttman, L., 1950, The principal components of scale analysis. In S. S. Stouffer (Ed.), Measurement and prediction (p. 312–361). Princeton NJ: University Press.
14) Hambleton, R. K. & Rovinelli, R. J., 1986, Assessing the dimensionality of a set of test items. Applied Psychological Measurement, 10, 287–302.
15) Hattie, J. A., 1984, An empirical study of various indices for determining unidimensionality. Multivariate Behavioral Research, 19, 49–78.
16) Hattie, J. A., 1985, Methodology review: Assessing unidimensionality of tests and items. Applied Psychological Measurement, 9, 139–164.
17) Holland, P.W., 1981, When are item response models consistent with observed data?. Psychometrika, 46, 79–92.
18) Holland, P.W. & Rosenbaum, P. R., 1986, Conditional association and unidimensionality in monotone latent variable models. Annals of Statistics, 14, 1523–1543.
19) Hulin, C. L., Drasgow, F. & Parsons, C., 1983, Item response theory: Applications to psychological measurement. Homewood IL: Dow & Jones Irwin.
20) Humphreys, L. G. & Ilgen, D. R., 1969, Note on a criterion for the number of common factors. Educational and Psychological Measurement, 29, 571–578.

21) Humphreys, L. G. & Montanelli, R. G., 1975, An investigation of the parallel analysis criterion for determining the number of common factors. Multivariate Behavioral Research, 10, 193-205.

22) 池田 央, 1973, テストII(心理学研究法, 7). 東京大学出版会.

23) Jöreskog, K. G., 1978, Structural analysis of covariance and correlation matrices. Psychometrika, 43, 443-447.

24) Kaiser, H. F., 1970, A second generation little jiffy. Psychometrika, 35, 401-415.

25) Krus, D. J. & Bart, W. M., 1974, An ordering theoretic method of multidimensional scaling of items. Educational and Psychological Measurement, 34, 525-535.

26) Krus, D. J. & Krus, P. H., 1980, Dimensionality of hierarchical and proximal data structures. Applied Psychological Measurement, 4, 313-321.

27) Linn, R. L., 1968, A Monte Carlo approach to the number of factors problem. Psychometrika, 33, 37-72.

28) Loevinger, J., 1948, The technique of homogeneous tests. Psychological Bulletin, 45, 507-529.

29) Lord, F. M., 1953, The relations of test score to the trait underlying the test. Educational and Psychological Measurement, 13, 517-548.

30) Lord, F. M. & Novick, M. R., 1968, Statistical theories of mental test scores. Reading MA : Addison-Wesley.

31) Lord, F. M., 1980, Applications of item response theory to practical testing problems. New York : Erlbaum Associates.

32) Lumsden, J., 1961, The construction of unidimensional tests. Psychological Bulletin, 58, 122-131.

33) McDonald, R. P., 1967, Non-linear factor analysis. Psychometric Monograph (No. 15)

34) McDonald, R. P., 1981, The dimensionality of tests and items. British Journal of Mathematical and Statistical Psychology, 34, 100-117.

35) McDonald, R. P., 1982, Linear versus nonlinear models in item response theory. Applied Psychological Measurement, 6, 379-396.

36) Muthen, B., 1978, Contributions to factor analysis of dichotomous variables. Psychometrika, 43, 551-560.

37) Reckase, M. D., 1979, Unifactor latent trait models applied to multifactor tests : Results and implications. Journal of Educational Statistics, 4, 207-230.

38) Reynolds, T., 1981, ERGO : A new approach to multidimensional item analysis. Educational and Psychological Measurement, 41, 643-659.

39) Rosenbaum, P. R., 1984, Testing the conditional independence and monotonicity assumptions of item response theory. Psychometrika, 49, 425-435.

40) Stout, W., 1987, A nonparametric approach for assessing latent trait unidimensionality, Psychometrika, 52, 589-617.

41) Tucker, L. R. & Lewis, C., 1973, A reliability coefficient for maximum likelihood factor analysis. Psychometrika, 38, 1-10.

42) van den Wollenberg, A. L., 1982, Two new test statistics for the Rasch model. Psychometrika, 47, 123-140.

43) Wise, S. L., 1983, Comparison of order analysis and factor analysis in assessing the dimensionality of binary data. Applied Psychological Measurement, 7, 311-321.

44) Wise, S. L. & Tatsuoka, M. M., 1986, Assessing the dimensionality of dichotomous data using modified order analysis. Educational and Psychological Measurement, 46, 295-301.

45) Wright, B. D. & Panchapakesan, N., 1969, A procedure for sample-free item analysis. Educational and Psychological Measurement, 29, 23-48.