態度テストの信頼性・妥当性・一次元性について

一次元性と信頼性の関係を中心に

小　林　久　高

1．はじめに

テストの構成に際して，信頼性と妥当性は絶対の必要条件だといわれる。信頼性はテストの精度に関わる性質である。また，妥当性は，測定者が測定しようとしている対象の性質を，そのテストが測定しているのかどうかということに関わっている性質である。信頼性と妥当性は区別される概念であり，相互に交換可能な概念ではない。例えば，ある人が気温を計ろうとしているとき，誤って湿度計を用いたならば，彼の測定の妥当性は低い。しかしその湿度計が正確に湿度を計るものならば信頼性は高いのである[1)]。

さて，本稿の対象とする態度テストは，通常，複数の項目を含んでいる。このようなテストの構成に際しては，信頼性と妥当性以外に一次元性といわれる性質もまた重視される。一次元性は，内的一貫性あるいは等質性・同質性とよばれることもあり，テストに含まれる各項目が一次元上に並ぶという性質を指している。すなわち，一次元性とは，テストに含まれるすべての項目間の相関が1であるという性質を指しているのである。したがって，相関が1に近づくほど，そのテストの一次元性の程度は強いといえる[2)]。

ところで，一次元性は，信頼性や妥当性といかなる関係にあるのか。この関係について一口で説明することはむずかしい。というのは論者によって，この関係は異なってとらえられているからである。社会学や社会心理学の比較的入門的な教科書においても，一次元性のとらえ方はさまざまであり，信頼性に含まれる1つの特徴あるいは信頼性そのものとみなす者もいれば（野口, 1985:78; 安藤, 1987:157），妥当性に関わる性質として考えられてきた概念であるとする者もいる（安田, 1970:131)。また一次元性を，信頼性・妥当性とは一応異なったテストの要件とみなす者もいる（安田, 1970:146;池田, 1971:131)。

本稿では，信頼性・妥当性の基本的意味およびその推定方法を明確にすることによって，信頼性と一次元性の関係，および妥当性と一次元性の関係を明らかにし，テストに必要な性質とは何かということについて議論してゆく[3)]。

2．妥当性と一次元性

すでに述べたように，妥当性の基本的な意味は，テストが測定者の計ろうとしている対象の性質を誤らずに計っているかどうかということに関わっている。ところで，信頼性が，測定者の測定意図とは関わらないいわば「テスト自体の性質」であるのに対して，妥当性は「テストと測定者の

1）ただし，信頼性と妥当性とは無関係ではない。妥当性係数（後述）は，信頼性係数（後述）の平方根を越えることはないのである。この関係については，Carmine (1979＝1983:32-33)，池田 (1980:82-83) などを参照されたい。

2）等質性・同質性は，単に一次元であるだけでなく，各項目の強さが同じであるという意味を含んでいる場合がある。すなわち，同一の被験者ならば，異なる項目でもおなじ得点をとるというのが等質性・同質性というわけである。そのように考えても，等質性・同質性はともに一次元性に含まれるので（等質性・同質性⊂一次元性），以下の議論に根本的な修正を加える必要はない。ただし，以下では等質性・同質性を一次元性と等価であると考えて議論を進める。

3）本稿の課題は，従来よりあった信頼性と妥当性の議論を，一次元性との関連で整理することにあるので，以下では数学的な証明の多くを引用文献にゆずる。

測定しようとする対象の間の性質」であることは忘れてはならない。もっとも，テストには通常，そのテストが何を計ろうとしているのかということに関する説明が付随しているので，その説明まで含めてテストと考えるならば，妥当性をテストの性質と見なしても不都合はない。

さて，テストの妥当性と一次元性の関係については，安田（1970:129-147）が明快な議論をしている。われわれは以下，彼にしたがって論を進めることにしよう。

テストの妥当性評価のためには，予測的尺度，内的尺度という2つの尺度を区別して考える必要がある。ここで予測的尺度とは，外部にすでに存在する定性的または定量的なある標識（外的基準）を予測することを目的とする尺度であり，内的尺度とは，観念的には考えられるが，何ら客観的な外在する基準の存在しない量的特性を測定する尺度をいう。そしてテストは，この予測的尺度か内的尺度かのどちらかなのである。

さて，テストが予測的尺度である場合，そのテストの妥当性を明らかにするのは簡単である。このような場合，通常，外的基準とテストの得点の相関が妥当性係数とされる。しかし，テストが内的尺度である場合，そのテストの妥当性はどのように評価すればいいのか。すなわち，ここで対象としているような態度テストの妥当性はどのように評価すればよいのだろうか。このようなとき，しばしば依りどころとされてきたのが，内的一貫性（一次元性）という基準であると安田はいう。この考え方からすると，内的尺度の妥当性とは，その一次元性を意味することになり，妥当性の高い態度テストとは，一次元性の強いテストということになる。

安田の議論は，そこからいくつかの一次元性を確保するための技法に向かい，サーストンの等現間隔尺度，リッカート尺度，項目分析，因子分析，ガットマンのスケイログラム分析などを説明してゆく。しかし，テストの妥当性に関して，彼が最終的に下す結論は，尺度の一次元性が確保されても「それが何を計っているのか，計るべきあるものを確かに計っているのか」という本来の意味での妥当性の問題は依然として残るというものであり，内的尺度の妥当性は，結局常識的判断と，プラグマティックな試行錯誤の過程によってのみ解決されるというものである。

以上，安田（1970）にしたがって，妥当性と一次元性について述べてきた。まとめておこう。慣習的な尺度理論においては，内的尺度の妥当性を評価するための代替的基準として，一次元性が用いられてきたが，一次元性の問題は妥当性の問題とは区別されるべきものなのである[4]。

3．信頼性と一次元性

信頼性と一次元性の関係は，妥当性と一次元性の関係に比べて，やや込み入っている。この込み入った問題を解きほぐしてゆくために，われわれは2つの水準を区別しなければならない。第1の水準とは信頼性自体の水準であり，第2の水準とは信頼性の推定値の水準である。2つの水準を混同すると，議論は非常にややこしいものになる。われわれは，3.1と3.2で信頼性自体の水準の議論をし，3.3と3.4でその推定について述べる。

3.1　信頼性の意味

テストの信頼性の意味をさらに明確にするために，まず，古典的テスト理論に基づいて，測定について述べておくことにしよう[5]。測定と誤差についての最も基本的な関係は，次の式で表現され

4）内的尺度の妥当性と一次元性との基本的関係については，安田の説明で十分である。ただし細かい点を述べると，次のようにいわなければならない。内的尺度の妥当性を検討する際に，別のテストとの相関をとり，それを併存的妥当性（⊂基準関連妥当性）と称することがある。このような妥当性を基準関連妥当性とよんでよいかという問題はさておき，この算出された値は，信頼性と1）で述べた関係にある。後述するように，信頼性は一次元性と関係しているので，結局，この値は一次元性と関連をもつことになるのである。

5）古典的テスト理論の概略については，Carmine & Zeller（1979），池田（1973）などを参照されたい。文献によっては，以下の4つの仮定式と異なる仮定をおいているものもあるが，実質的な意味は変わらない。最近のテスト理論は，この古典的な立場よりも進んだものである（Item Responce Theory）。しかし，現在でも社会学や社会心理学の一般的なテスト作成の原理は，この古典的なテスト理論であるので，本稿でも古典的な考え方のみに焦点を当てて考察を進めることにする。

る[6]。

$$X = T + e \quad (1)$$

X：測定値
T：真値
e：測定誤差

つまり，ある人のある性質を測定する場合，測定された値は，彼のその性質についての真値と測定誤差の和として表現される。この測定値，真値，測定誤差の関係式は，単一の項目しか含まないテストの得点についての式としてもとらえられるし，複数の項目を含むテストの総合得点についての式としてもとらえられる。

さて，上述の関係式(1)は，特定の個人についての関係式であることに注意すべきである。今問題にしているのは，複数の人びとを含む集団に対して適用される態度テストの信頼性であるので，議論をその方向に進めよう。

まず，被験者の集団に対してテストを1回行った場合を想定して，次の仮定をおくことにする。

仮定1　測定誤差の期待値は0である。

$$E(e) = 0 \quad (2)$$

仮定2　真値と測定誤差の相関は0である。

$$r(T, e) = 0 \quad (3)$$

次に，同一の被験者の集団に対して2回テストが行われた場合を想定して，次の2つの仮定をおく。

仮定3　ある測定（テスト1）における測定誤差と別の測定（テスト2）における測定誤差の相関は0である。

$$r(e_1, e_2) = 0 \quad (4)$$

仮定4　ある測定（テスト1）における測定誤差と，別の測定（テスト2）における真値の相関は0である。

$$r(e_1, T_2) = 0 \quad (5)$$

以上，4つの仮定はさほど無理な仮定ではないので，以後の議論はこれらの仮定がすべて成り立っていることにして進めよう。これらの仮定を用いると（正しくは仮定1と仮定2を用いると），次の式が成立する。

$$var(X) = var(T) + var(e) \quad (6)$$

すなわち，測定値の分散は，真値の分散と測定誤差の分散の和である。

さて，テストの精度を示す信頼性係数（ρ）は，この(6)式をもとに次のように表現される。

$$\rho = var(T)/var(X) \quad (7)$$

すなわち，信頼性係数は測定値の分散に占める真値の分散の割合である。

複数の項目を含む態度テストの信頼性は，総合得点の測定値の分散に占める総合得点の真値の分散の割合として定義される。つまりテスト全体で1つの測定と考えるのである。

信頼性の測度としては，信頼性係数以外にも，測定の標準誤差，信頼性指数，SN比などがある。したがって，信頼性係数とは信頼性を計る1つの測度にすぎない。しかし，普通，信頼性というとき，この信頼性係数を指すことが多いので，本稿でも，以下，そのように用いることにする。

3.2　信頼性と一次元性の関係

さて，本稿の主目的は，信頼性係数について説明することではなく，信頼性と一次元性との関係について述べることであった。これまでの信頼性についての議論から，一次元性と信頼性の関係を考えるとき，両者は原理的には関係しないという結論が導けそうに見える。しかし，それは正確で

6）式(1)〜(4)は，古典的テスト理論の最も基本的な式であるが，系統的な誤差の影響を無視しているので完全なものとはいえない。この系統誤差を考慮に入れた考え方も存在するが，本稿では，それを考慮に入れない最も基本的な考え方を中心に議論する。

はない。

いま，N個の項目（...i,j,..）を含むテストがあるとき，一般に次の式が成り立つ[7]。

$$\rho=[\beta'+(N-1)\gamma']/[1+(N-1)\gamma'] \quad (8)$$

ρ ：テストの信頼性係数

β' ：各項目測定値の分散平均に占める各項目真値の分散平均の割合。すなわち，

$\beta'=\mu[var(T_i)]/\mu[var(X_i)]$

γ' ：各項目測定値の分散平均に占める各項目測定値間の共分散平均の割合。すなわち，

$\gamma'=\mu[cov(X_i, X_j)]/\mu[var(X_i)]$　$i\neq j$

ここで，各項目得点が標準得点化され，分散が等しいと仮定するならば，より単純な次の関係が成り立つ。

$$\rho=[\beta+(N-1)\gamma]/[1+(N-1)\gamma] \quad (9)$$

ρ：テストの信頼性係数

β：各項目の信頼性係数の平均

γ：測定値をもとにした項目間相関係数の平均（自身との相関は除く）

式(9)を見ればわかるように，複数の項目を含むテストの信頼性は，そこに含まれている各項目の信頼性だけでなく，項目間相関や項目数にも影響されるのである。表1は，項目得点が標準得点化され，それらの分散が等しいとき，各項目の信頼性係数の平均を.8とすると，項目間相関の平均と項目数とによって，テストの信頼性係数がどのような影響を受けるのかを明らかにしたものである。

表より，項目数が多くなればなるほど，項目間の相関が高くなればなるほど，テスト全体の信頼性が増加することが読み取れよう。このように，相関の高い項目がテストに含まれることによってテスト全体の信頼性が高まる効果を，ここでは合成の効果ということにしよう。

ところで，表1の第1列（項目間相関が.0の列）には注意すべきである。この列は，相互に全く相関しない項目を含むテスト，つまり項目が決して一次元的ではないテストを表している。このようなテストにおいては合成の効果は見られない。しかし，各項目の信頼性がかなり高いならば，テスト全体の信頼性も高くなるのである。つまり項目間の一次元性や項目数の多さは，合成の効果の必要条件であるが，テストが高い信頼性をもつことの必要条件ではない。

表1　項目数と項目相関のテストの信頼性への影響

項目数	項目間相関の平均（γ）				
	.0	.2	.4	.6	.8
5	.800	.889	.923	.941	.952
10	.800	.929	.957	.969	.976
15	.800	.947	.970	.979	.984
20	.800	.958	.977	.984	.988
25	.800	.966	.981	.987	.990
30	.800	.971	.984	.989	.992
35	.800	.974	.986	.991	.993
40	.800	.977	.988	.992	.994
45	.800	.980	.989	.993	.994
50	.800	.981	.990	.993	.995

項目の信頼性係数＝.8の場合[8]

まとめておこう。項目数が一定のとき，項目間の一次元性の程度を強めることは，テストの信頼性の増加に役立つ。しかし，テストの信頼性が高いからといって，そこに含まれる項目が一次元的であるとはいえない。項目の一次元性はなくとも，十分信頼性の高いテストは存在するのである。すなわち，一次元性を高めることは，信頼性を高めるための十分条件であるが，必要条件ではないのである。

さて，信頼性と一次元性の関係が，今述べた信頼性係数の水準だけで登場するのであるならば，話はさほど複雑ではない。問題を混乱させているのは，一次元性の問題が，信頼性係数の推定という水準の議論にも関わっていることによる。われわれは次に，この新たな水準の議論に向かうことにしよう。しかしその前に推定の前提となる平行測定について述べておかねばならない。

3.3　平行測定と一次元性

式(7)で定義された信頼性係数を，直接算出する

7）式(8)(9)の証明は，池田（1980:84-86）参照のこと。

8）項目間相関の平均は項目の信頼性係数の平均を越えない（池田, 1980:92）。

ことは不可能である。なぜなら，真値の分散は，実際にはとらえられないからである。そこで，平行測定という考え方が導入される。ここでは，平行測定だけでなくそれに関連するいくつかの測定，およびそれらと一次元性との関連について述べる[9]。

以下の議論でも，測定を項目による測定と考えてもよいし，複数の項目を含んだテストによる測定と考えてもよい。前者の場合は項目得点について議論しており，後者の場合は総合得点について議論していることになる。この項での議論は，どちらを考えても成り立つ。測定は複数の被験者よりなる集団を対象としていることにも注意されたい。

まず，2つの測定があると考える。X_{i1}，X_{i2} をそれぞれの測定における被験者(i)の測定値，T_{i1}，T_{i2} をそれぞれの測定における被験者(i)の真値，e_{i1}，e_{i2} をそれぞれの測定における測定誤差とするとき，この2つの測定は次の式で表現される。

$$X_{i1}=T_{i1}+e_{i1} \tag{10}$$
$$X_{i2}=T_{i2}+e_{i2} \tag{11}$$

（1）平行測定

平行測定とは，すべての被験者(i)について，次の式が成り立つような2つの測定を意味する。

$$T_{i1}=T_{i2} \text{ かつ } \mathrm{var}(e_1)=\mathrm{var}(e_2) \tag{12}$$

平行測定においては，両測定は同じ性質を同じ強さ（同じ目盛間隔）で計っており，その精度も等しい。また平行測定では次の5つの関係が成立する。

$$\mu(T_1)=\mu(T_2) \tag{13}$$
$$\mu(X_1)=\mu(X_2) \tag{14}$$
$$\mathrm{var}(T_1)=\mathrm{var}(T_2) \tag{15}$$
$$\mathrm{var}(X_1)=\mathrm{var}(X_2) \tag{16}$$
$$r(T_1, T_2)=1 \tag{17}$$

（2）τ 等価測定

平行測定よりも条件のゆるい測定として，τ 等価測定がある。すべての被験者(i)について，2つの測定間で次の式が成立する場合，それらは τ 等価測定であるという。

$$T_{i1}=T_{i2} \tag{18}$$

τ 等価測定においても，両測定は同じ性質を同じ強さで計っているといえる。しかし，その精度は必ずしも等しいわけではない。

平行測定のところで述べた5つの関係のうち，τ 等価測定においても成り立つのは次の関係である。

$$\mu(T_1)=\mu(T_2) \tag{19}$$
$$\mu(X_1)=\mu(X_2) \tag{20}$$
$$\mathrm{var}(T_1)=\mathrm{var}(T_2) \tag{21}$$
$$r(T_1, T_2)=1 \tag{22}$$

（3）本質的 τ 等価測定

τ 等価測定よりも少し条件のゆるい測定として，本質的 τ 等価測定がある。すべての被験者(i)について，2つの測定間に次の式が成立するとき，それらを本質的 τ 等価測定という。

$$T_{i2}=T_{i1}+c_{12} \tag{23}$$

これは，測定の原点を（定数 c_{12} を加えることによって）移動させれば，両測定が同じ性質を同じ強さで計っているといえることを意味している。測定の精度は必ずしも等しくはない。

平行測定のところで述べた5つの関係のうち，本質的 τ 等価測定においても成り立つのは次の関係である。

$$\mathrm{var}(T_1)=\mathrm{var}(T_2) \tag{24}$$
$$r(T_1, T_2)=1 \tag{25}$$

（4）一次元的測定

平行測定について通常議論されるのは，上述の平行測定，τ 等価測定，本質的 τ 等価測定の3つ

9）平行測定，τ 等価測定，本質的 τ 等価測定については，Novick（1967）の議論が簡潔で要領を得ている。

であるが，本稿は一次元性を問題にしているので，ここで一次元的測定というものを考えておこう。

一般に，特性Zと特性Yに一次元的な関係があるとき，定数a，bを用いた次の式が成り立つ。

$$Y=aZ+b \tag{26}$$

一次元的測定とは，この関係を基にした測定である。すなわち，すべての被験者(i)について，次の式が成立するとき，両測定は一次元的であるという。

$$T_{i2}=a_{12}T_{i1}+b_{12} \tag{27}$$

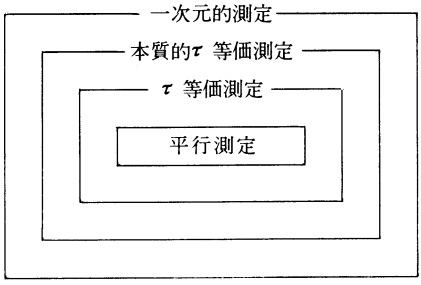
一次元的測定においても，両測定はおなじ性質を計っているといえる。しかし原点と計る強さ（目盛の原点と間隔）は両測定で異なっている可能性がある。また両測定の精度も必ずしも同じではない。

平行測定のところで述べた5つの関係のうち，一次元的測定においても成り立つのは次の関係だけである。

$$r(T_1, T_2)=1 \tag{28}$$

以上，平行測定に関連した4つの測定について述べてきた。最後にそれら相互の関係をまとめると図1が得られる。

図1　平行測定とそれに関連した測定の関係



3.4　信頼性係数の推定法と一次元性

信頼性係数は，実際には直接算出できないので，これまで述べてきた平行測定やそれに関連する測定にもとづいて推定される。このとき，推定値が真の信頼性係数と等しくなるとは限らない。推定値はある条件が満たされているときにのみ，真の信頼性係数と等しくなるのである。

以下，いくつかの代表的な信頼性係数の推定法を紹介するとともに，①その方法で得た信頼性推定値が真の信頼性と等しくなるための条件（ただし上述の4つの仮定以外），②その条件とテストの各項目真値の一次元性との関係，③その方法によって得られる信頼性推定値と，測定値の一次元性との関係，④正しい推定の条件が満たされていないとき，その方法で得られた信頼性推定値は，真の信頼性といかなる関係にあるか，という4点について述べる。以下でも，多項目を含む態度テストを中心に議論する。

（1）再テスト法

再テスト法とは，同一の被験者集団に対して，2回，同じテストを行い，その相関係数を信頼性係数の推定値とするという方法である。

①信頼性推定条件

この方法によって求められた信頼性推定値が真の信頼性である条件は，1回目のテストと2回目のテストが平行測定であることである（池田，1973:138)。両テストが平行測定といえるか，ということに関しては，いくつかの疑問が提出されている。その中でしばしば問題になるのは，記憶によって両テストの独立性が保たれないというものである。しかし，態度テストの場合，学習の効果よりも態度そのものの変化が問題になろう。異なった時期に行われる2つのテストの間で，被験者の真の態度に変化があるならば，「両テストの真値が等しいこと」という平行測定の前提が破壊されてしまうのである。

②信頼性推定条件と真値の一次元性

再テスト法の条件である平行測定と一次元性との間には，図1のような関係が存在することはすでに述べた。したがって，両テストの真値の総合得点間の関係は一次元的でなければならない。しかし，テスト内の各項目の真値は，一次元的である必要はないのである。したがって，2回のテストが平行測定であるという条件さえ整えば，再テスト法によって，一次元的であるとみなされない複数項目を含むテストの信頼性係数を正しく推定

することができる。

③信頼性推定値と測定値の一次元性

3.2で述べた影響（合成の効果→真の信頼性の増大→信頼性推定値の増大）を除いて，テスト内の各項目測定値間の相関も，再テスト法による信頼性係数の推定には関連しない。

④正しい推定条件が満たされていない場合

2回の測定が平行測定になっているかどうかは，実質的には知り得ない問題である。正しい推定条件が満たされていない場合，再テスト法による信頼性係数の推定値と真の信頼性係数との間に，明確な関係はない。

（2）代替テスト法

代替テスト法とは，あるテストの代替的なテストを作り，もとのテストと代替的なテストを，ともに同一の被験者集団に対して実施し，両テストの相関係数をとり，それを信頼性係数の推定値とするという方法である。

代替テスト法は多くの点で再テスト法に似ているので簡単にその性質を述べよう。

①信頼性推定条件

2つのテストが平行測定であること（池田，1973:139)。

②信頼性推定条件と真値の一次元性

テスト内の各項目真値の一次元性は必要条件として設定されていない。

③信頼性推定値と測定値の一次元性

テスト内の各項目測定値間の相関も，代替テスト法による信頼性係数の推定には実質的に関連しない（ただし3.2で述べた影響を除く)。

④正しい推定条件が満たされていない場合

推定条件が満たされていない場合，信頼性係数の推定値と真の信頼性係数との間に明確な関係はない。

（3）折半法

折半法は，テストの全項目を半分ずつ2つに分け，両部分テストの相関を算出し，その相関係数を何らかの式で修正し，信頼性係数の推定値を導くという方法である。ここでは，折半法の中で最もよく行われる，スピアマン＝ブラウンの修正公式を利用した信頼性推定について述べる[10]。

スピアマン＝ブラウンの修正公式を用いると，信頼性係数の推定値は次の式で表せる。

$$\rho' = 2r(X_1, X_2)/[1 + r(X_1, X_2)] \quad (29)$$

ここで，ρ'はテスト全体の信頼性係数の推定値を表し，$r(X_1, X_2)$は折半したテスト間の相関を表している。

①信頼性推定条件

スピアマン＝ブラウンの公式を用いた折半法の信頼性係数推定値が，真の信頼性係数と等しくなる条件は，折半された部分テスト双方が平行測定であることである（池田，1973:141)。この条件が満たされたときにのみ，推定値は真の信頼性係数と等しくなる。

②信頼性推定条件と真値の一次元性

この推定法も，テスト内の各項目真値の一次元性を必要条件とするものではない。ただし，部分テストは相互に平行測定でなければならないので，部分テストの総合得点の真値間の関係は，一次元的でなければならないということになる。

③信頼性推定値と測定値の一次元性

式(29)を見ればわかるように，この推定値は，部分テストの測定値の相関によって一義的に定まる。したがって，両部分テストの測定値の総合得点が一次元的な関係にあるほど，この推定値は大きくなる。

しかし，この信頼性推定値が大きくなるために，折半する前のテストの各項目の一次元性（測定値）が必要とは限らない。ただし，各項目が一次元的（測定値）であることは，両部分テストが一次元的（測定値）であることの十分条件であるし，3.2で述べた関係も存在するので，そのような場合には信頼性推定値は大きくなると予測できる。

④正しい推定条件が満たされていない場合

推定条件が満たされていない場合，この方法による信頼性係数の推定値と真の信頼性係数との間に明確な関係はない。

（4）内的一貫性法

次に，内的一貫性法という推定法について述べ

10）折半法には他にもいろいろな公式を用いたものがある。池田（1973:142-143）はそれらを紹介している。

よう。信頼性と一次元性の関係をややこしくしている原因の1つは，この内的一貫性法と命名された信頼性係数の推定法が存在するということにある。その推定法は，名前からもわかるように，深く一次元性と関連している。そしてその関連の仕方は，少々複雑なのである。

内的一貫性法をとる信頼性係数の推定値のうち，最もよく利用されるのは，クロンバックのα係数と，クーダー＝リチャードソンのKR20という係数である。

αやKR20は，どのように内的一貫性（一次元性）と関わっているのだろうか。入門書においてしばしばいわれることは，αやKR20がテストの諸項目の等質性（内的一貫性）を計っているということである（安田, 1969:286;野口, 1985:78;安藤, 1987:157)。そこから議論は二手に分かれる。一方は信頼性とは等質性のことであるかのようにいい（野口, 1985;安藤, 1987)，他方は，αやKR20では信頼性を計っていることにならないので，再テスト法を用いなければならないとするのである（安田, 1969)[11]。

では，いずれの主張が正しいのだろうか。あるいは，いずれの主張とも正確とはいえないのかもしれない。一体，内的一貫性法と命名されたαやKR20は，信頼性や一次元性とどう関連するのだろうか。以下，この点について明らかにしていくことにしよう。KR20はαの特殊形態と考えられるので，αを中心に議論する。

クロンバックのαは，次の式で表される。

$$\alpha=[N/(N-1)][1-\Sigma var(Y_i)/var(X)] \quad (30)$$

N　　：項目数
var(X)　：テストの得点（合成得点）の分散
$\Sigma var(Y_i)$：テスト内の各項目(Y_i)の分散の合計

各項目得点が標準得点化され，分散が等しいときには，αは，項目間相関の平均γを用いて次のようにも表現できる。

$$\alpha=N\gamma/[1+\gamma(N-1)] \quad (31)$$
$$\alpha=N\gamma/[N\gamma+(1-\gamma)] \quad (32)$$

①信頼性推定条件

αは信頼性係数と全く無関係ではない。それは特定の条件が満たされるならば，信頼性係数の正しい推定値となる。しかし，逆にいえば，特定の条件が満たされない限り，正しい信頼性係数の推定値とはならないのである。その条件とは，テストに含まれている諸項目が，本質的τ等価であることである（Novick, 1967:6-7)。

②信頼性推定条件と真値の一次元性

したがって，αが信頼性係数に等しくなるためには，テスト内の諸項目（真値）が一次元上に並ぶことが必要条件なのである（ただし十分条件ではない。なぜなら本質的τ等価⊂一次元だから)。この意味において，内的一貫性法の係数であるαは，一次元性と大いに関連している。この推定条件としての項目間の一次元性が意味するところは，次のようなことである。すなわち，各項目の真値が決して一次元的ではないと予測されるときに，αを用いて信頼性係数の推定値を出しても，その推定値は真の信頼性係数を正しく推定していないということなのである。よくあるα自身が信頼性係数であるという考え，αはつねに正しく信頼性係数を推定するという考えは，ともに誤っている。

③信頼性推定値と測定値の一次元性

αは測定値の内的一貫性の測度としても，用いられてきた（Novick, 1967:11-12)。式(32)を見れば，それもうなずける。γが大きくなればなるほどαは大きくなるからである。しかしこれにはNが一定であるという条件がある。つまり，Nが大きくなってもαは大きくなるのである。ここでαとNとγの関係を明らかにしよう（表2)。

通常の態度テストには，5〜50程度の項目が含まれているが，20項目程度のテストの場合，平均相関が0.2程度でも$\alpha=0.83$となってしまうことが表2よりわかる。完全な一次元性が成立するとき，$\gamma=1$であることを考慮すると，αで測定値の

11) 安田（1969）のこの立場は，林（1950）に影響を受けたものと思われる。なお，『社会調査ハンドブック（第3版)』においても，安田の立場に変わりはない。

表2　項目数と項目相関のαへの影響

項目数	項目間相関の平均（γ）					
	.0	.2	.4	.6	.8	1.0
5	.000	.556	.769	.882	.952	1.000
10	.000	.714	.870	.938	.976	1.000
15	.000	.789	.909	.957	.984	1.000
20	.000	.833	.930	.968	.988	1.000
25	.000	.862	.943	.974	.990	1.000
30	.000	.882	.952	.978	.992	1.000
35	.000	.897	.959	.981	.993	1.000
40	.000	.909	.964	.984	.994	1.000
45	.000	.918	.968	.985	.994	1.000
50	.000	.926	.971	.987	.995	1.000

Carmine & Zeller, 1978＝1983：43をもとに作成
項目得点が標準化され分散が等しい場合

一次元性を計ることの危険性は目に見えている。さらにいえば，測定値の一次元性の測度としてわざわざ複雑なαを用いなくても，より単純なγを用いれば十分なのである。

④正しい推定条件が満たされていない場合

ところで，一般にαは信頼性係数（ρ）とつねに一定の関係にあることが知られている。すなわち，

$$\rho \geqq \alpha \qquad (33)$$

がつねに成り立っているのである。等号は，テスト内の諸項目が本質的τ等価測定であるとき成立し，本質的τ等価という条件が満たされていないとき，αは真の信頼性係数よりつねに小さな値をとる（Novick, 1967:4-6）。この関係が存在する意義は大きい。テストの各項目の真値間の関係は，実際は知り得ないものなので，そのテストが本質的τ等価かどうかも通常知ることができない。ゆえに推定には危険がつきまとう。しかしαを用いるならば，少なくとも信頼性係数を過大に評価する危険はない。この性質ゆえにαは信頼性係数を推定する非常に有効な方法となるのである[12]。

しかし，この点を重視して，αこそ信頼性を推定する決定的な係数であると考えるのは危険である。もう一度，表1と表2を見てみよう。これらはそれぞれ，項目数と項目間相関平均を異にする多数のテストの，①真の信頼性係数（表1）と，②その推定値であるα（表2）を表しているとみなすことができる。2つの表より，今述べた$\rho \geqq \alpha$が成立していることがわかる。しかし，ここで注意すべきことは別のことである。すなわち，テストの真の信頼性係数が .8 であっても，項目の相関が .0 であるならば，信頼性推定値αは .0 となってしまうということである。信頼性という考え方自体は，そもそも一次元性を意味しないし前提ともしない。すなわち，必要十分条件でもないし必要条件でもない。にもかかわらず，αを用いた信頼性推定は，そこに実際に存在する以上の関連を誤って連想させるのである。

以上の説明から，α（そしてその特殊形態としてのKR20）と一次元性との関係が明らかになった。αはやはり信頼性係数の推定値であり，その正しい推定の条件が，テスト内の諸項目で計られる真値の一次元性（正しくは本質的τ等価性）なのである。また，αが測定値の一次元性（内的一貫性）を計っていると考えるのは正確でないのである。したがって，内的一貫性法という命名は，その係数の計る対象ではなく，その係数が信頼性を正しく推定するための条件に由来すると考える方が，誤解が少ないと思われる。

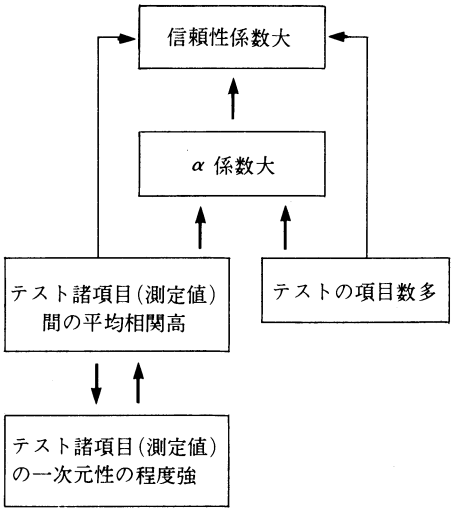
ところで，よく，信頼性を高めるために相関の高い項目を付け加えるということがいわれるが，そのことは2つの効果をもっていることが今までの説明から理解されよう。第1の効果は，そのことによって真の信頼性係数の値そのものが高められる（合成の効果）ということである（3.2参照）。第2の効果は，信頼性係数の推定値であるαの値が高められるというものである。2つの効果は区別して考える必要がある。

最後に，信頼性係数・α・測定値の一次元性の論理的な関係について，図にまとめておこう（図2）。図の矢印の方向が示しているように，真の信頼性係数の値が大きくても，αの値が必ず大きくなるわけではない[13]。

12）しかし，α係数より良い信頼性の下限値を与える推定値も存在する（池田, 1973:85-87）。

13）以上の説明で，本項(4)の最初に述べた議論の混乱も解決可能であろう。混乱の最も大きな原因は，双方とも信頼性の水準と信頼性推定の水準を明確に区別しないことにある。

図2　信頼性・α・一次元性の論理的関係



4．おわりに

以上，一次元性と妥当性，信頼性の関係について述べてきた。要点をまとめておこう。以下，テストは複数の項目を含み，一次元性とは各項目間の測定値の一次元的関係を意味する。真値の一次元性は「一次元性（真値）」と記述する。

（1）テストの一次元性はテストの妥当性とは区別される概念である。

（2）テストの一次元性はテストの信頼性と区別される概念であり，信頼性と一次元性は等価でない。

（3）しかし，一次元性と信頼性の間には密接な関係がある。すなわち，テストの項目数が同じであり，そこに含まれている項目の信頼性も同じであるならば，一次元性の程度のつよいテストほど信頼性も高くなる。しかし，逆は真でない。完全に一次元的でないテスト（たとえば項目間相関が0のテスト）も高い信頼性をもち得る。つまり，一次元性を強めることは，信頼性を高めることの十分条件であるが，必要条件ではない。

（4）一次元性は，信頼性だけでなく信頼性の推定法にも関わっている。特にクロンバックのαは，テストの一次元性（真値）を必要条件とした信頼性の推定値であり，一次元性（真値）が確保されないときには正しい推定値にはならない。

（5）クロンバックのαは，信頼性係数の下限値である。しかし，一次元的でないテストの信頼性を過小評価する可能性をもつ。

（6）クロンバックのαは一次元性（内的一貫性）を正確に測定するものではない。一次元性はαによる推定の条件であり，αの測定する対象ではないのである。

最後に，一次元性に関連した2つの問題について述べ議論を終えることにしよう。

第1の点はテストの要件に関わる。われわれははじめに，複数項目を含むテストの構成の際，信頼性・妥当性・一次元性が重視されると述べた。そして一次元性と信頼性・妥当性の関係が曖昧であるとしたのである。ところで，一次元性はテストの絶対の必要条件であるとまで言い切れるのであろうか。なるほど多次元的な項目の得点を加えて，総合得点を算出するという操作の根拠は強固なものではない。しかしそれが無意味とまでいえるのかどうかは疑問が残る。

たとえば，F尺度についての因子分析的研究の多くが，その多次元性を指摘している（Krug, 1961;Kerlinger & Rokeach, 1966など）。ゆえにF尺度は無意味であるという説もある（Altemeyer, 1981：18-25, 112）。しかし，仮に，われわれが日常生活上，現実にそのような多元的な意味で「権威主義」という用語を用い，何らかの形で，総合得点を出し，権威主義の程度を云々しているとするならば，多次元的なテストの総合得点の算出も，すぐさま無意味と片付けるわけにはいかないだろう。われわれは現にある人びとの態度を計っているのであるから。この問題についてはこれ以上述べないが，意味論的な考察も加え，さらなる検討が必要とされよう[14,15]。

第2の点は，概念間の関連ということに関わる。今，2つの概念があるとしよう。それらの間の関係は，①分析的関係（概念的関係）と②経験的関係に区別される。ここで分析的関係とは，そ

14）この例の権威主義のように，相互に関連しない要素を内部にもつ概念については，ヴィトゲンシュタインの「家族類似性」に関する議論が参考になる（Wittgenstein, 1953＝1968:276-282）。

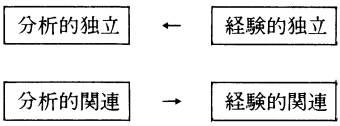
15）安田（1970:147）は，多次元的な内的尺度の構成法を考察すべき重要な問題としている。

の概念間の関係が論理内在的に導き出せる関係であり，経験的関係とは，概念間の関係が経験的な調査などで導き出せる関係である。

ところで，分析的関係と経験的関係に関しては，次の2点を注意すべきである。第1は，一方に関連があったからといって他方の関連をつねに導き出せるとは限らないという点である。たとえば，性別と髪の長さの間には分析的関連はない。すなわち，分析的に考えれば，それらは相互に独立である。しかし経験的な調査をすれば，多くの場合，その2つの概念（性別と髪の長さ）の間には関連があろう。したがって，われわれは分析的関係と経験的関係を区別して考える必要がある。

しかし，第2に，分析的関係と経験的関係との間には，2つの重要な関係があるというのも事実である。すなわち，①2概念が分析的に関連しているときには必ず，それらの概念間に経験的関連があり，②2概念が経験的に独立している場合には必ず，それらの概念は分析的にも独立しているということである。論理的な含意関係で考えるならば，分析的関係と経験的関係との間に存在するのは，以上2つ関係だけである（図3）。

図3　2概念の分析的関係と経験的関係



このことを念頭に置いておくと，因子分析などを利用する際に，混乱した議論を避けることができる。たとえば，F尺度を因子分析にかけると，いくつかの次元が抽出されよう。したがって，経験的に独立した概念がF尺度で計られているということができよう。そこから，（図3にしたがって）F尺度は，分析的に独立したいくつかの概念を計っているということもできる。ところで，その因子分析において，第一因子に，サド・マゾ的性格を表す諸項目と保守主義を表す諸項目とが高く負荷しているとしよう。すなわち，両者は経験的には関連しているのである。しかし，そこから両者は分析的にも関連した概念であると言い切ることはできない。図3が示しているように，経験的な関連は分析的関連を含意しないのであるから。したがって，それらの概念間に分析的な関係があるかどうかを検討するためには，さらなる概念的な検討が必要とされるのである[16]。

以上，分析的関係と経験的関係について述べた。テスト作成に関しては，経験的な関係だけでなく分析的関係にも考慮し，経験的関係と分析的関係の間柄について十分認識しておくことが重要といえるだろう。

引用文献

Altemeyer, B., 1981, *Right-Wing Authoritarianism*, University of Manitoba Press.

安藤清志 1987「態度・性格尺度の構成」末永俊郎編『社会心理学研究入門』東京大学出版会

Carmine, E. G., & Zeller, R. A., 1979, *Reliability and Validity Assesment*, Sage.（E. G.カーマイン・R. A.ツェラー『テストの信頼性と妥当性』水野欽司・野嶋栄一郎訳，朝倉書房，1983）

林知己夫 1951「所謂 RELIABILITY の測定について」『心理学研究』，21－2：61－64.

池田央 1971『行動科学の方法』東京大学出版会

池田央 1973『テストⅡ』東京大学出版会

池田央 1980『調査と方法』新曜社

Kerlinger, F., & Rokeach, M., 1966, "The Factorial Nature of the F and D Scale", *Journal of Personality and Social Psychology*, 4-4 : 391-399.

Krug, R. E., 1961, "An Analysis of the F Scale : 1. Item Factor Analysis", *The Journal of Social Psychology*, 53 : 285-291.

野口裕之 1985「テストをテストする」海保裕之編『心理・教育データの解析法10講基礎編』福村出版

Novick, M. R., 1967, "Coefficient Alpha and the Reliability of Composite Measurements", *Psychometrika*, 32-1 : 1-13.

Wittgenstein, L. 1953, *Philosophische Untersuchungen*（ヴィトゲンシュタイン「哲学探究（抄）」『論理哲学論考』藤本隆志・坂井秀寿訳，法政大学出版局，1968）

安田三郎 1969『社会調査ハンドブック（新版）』有斐閣

安田三郎 1970『社会調査の計画と解析』東京大学出版会

安田三郎・原純輔 1982『社会調査ハンドブック（第3版）』有斐閣

16）構成概念妥当性，中でも因子的妥当性を検討する際には，この点はとりわけ重要である。

付記

本稿の草稿に目を通し，最新のテスト理論に基づいて有益なコメントをくださった，関西学院大学社会学部立木茂男 講師に，深く感謝します。