

尺度構成法

—測れないものを測る方法について—

高木廣文・林 邦彦

尺度構成法

—測れないものを測る方法について—

高木廣文 TAKAGI Hirofumi
新潟大学医学部保健学科

林 邦彦 HAYASHI Kunihiro
群馬大学医学部保健学科

はじめに

身長や体重を測定するのは、日常的に当たり前のことだが、頭の善し悪しは本当に知能指数で測れるのだろうか。手術の前の患者の不安や恐れはどのくらいの大きさなのだろうか。このように、概念上では当然存在するはずなのだが、いざ測定しようとするときわめて難しいものが、現実にはたくさんある。たとえば、難病患者のQOLなどは、概念的には理解できても、実際にどのように測定したらよいのだろうか。

現実には何とかして測りたいのだが、それほど容易に測定するわけにはいかない事象は、どのように測ればよいのだろうか。ある事象を測定するための専用の「ものさし」を作成する方法は、「尺度構成法 (scaling)」として知られている。今回は、尺度構成法に関する主要な話題を簡単に説明しよう。

簡単に尺度を作るには

専用の「ものさし」は「尺度 (scale)」とよばれる。新しい尺度を作る手順は、実際にはそれほど難しいものではない。不安や共感性のように、測定のためのものさしがないものは、通常の測定方法では測ることはできない。しかし、そのような特性を測定するための手段として、比較的簡単でよく用いら

れる方法がある。すなわち、尺度を作りたい概念に関係すると考えられる項目を多数作成し、それぞれの項目に対する回答に適当な得点を与え、その総合得点でQOL、不安、共感性などを測定するという方法である。

①は、共感性尺度の一例を示したものである。ここでは、6つの質問項目があり、それぞれの質問に、「3. はい」、「2. どちらでもない」、「1. いいえ」のうちから一つだけ回答を選ぶようにしてもらおう。その後、各回答カテゴリに対して、3, 2, 1の得点を与え、その合計点で共感性を測定するものである。この方法は、「尺度化」、「尺度構成」などとよばれ、分野によっては得点 (スコア) を与えるので、「スコア化」などともよばれることがある。

このように、尺度を複数の項目に対する回答から得点化することで構成する場合、それらの項目の内

- 1) 他人の気持ちが人一倍よくわかる
- 2) たとえ自分は損をしても友人の苦境は見過ごせない
- 3) 他人の苦しみがよくわかる
- 4) 人のために尽くすことは好きだ
- 5) 人一倍感受性が豊かである
- 6) 捨て犬、捨て猫を見捨てるのはつらい

① 共感性尺度の例

容がその尺度の概念に適切なものでなければならぬことはいうまでもない。最も重要な点は、作成した尺度が自分が考えた概念を正しく測定しているのかということである。また、ものさしがものさしとして役に立っているのかも問題である。すなわち、朝と夕方でその測定結果が大きく異なるようでは、ものさしとして役に立つとはいいいくいだらう。

一般に、上記のような手順で尺度構成された測定用具の適切さは、「妥当性 (validity)」と「信頼性 (reliability)」の2つの側面から評価する必要がある。以下に、これらの考え方について簡単に説明しよう。

尺度構成での妥当性について

尺度の妥当性とは、「測定しようとしている概念内容を、どの程度適切に測定しているか」を示す用語である。作成した尺度の妥当性の評価は、以下に述べるようにいくつかのタイプに細分して考えることができる。

◆基準関連妥当性

大学入学試験の妥当性は、何によって評価できるだろうか。一つの考え方として、大学入学後の成績と入学試験の相関が高ければ、入学試験は妥当であったと考えることができるだろう。また、生活習慣を測定する場合、そのうちの一つとして、質問紙を用いて塩分の摂取量を推定したとしよう。推定した塩分摂取量と、実際の食事調査での塩分摂取量との相関が高ければ、同様に質問紙での塩分摂取量推定法は妥当性をもつと考えてよいだろう。

このように、新たな尺度の妥当性を調べるために、適当な評価基準（外的基準）との相関で検討される妥当性は「基準関連妥当性 (criterion-related validity)」とよばれている。

さらに、基準関連妥当性は、問題にしている尺度の測定と外的基準の測定時期によって2つに分類して

考えることができる。

質問紙調査による塩分摂取量推定法と栄養調査による食塩摂取量などのように、その測定が2つとも同時に実施されている場合、「同時的妥当性 (concurrent validity)」とよばれている。一方、大学入学試験と大学入学後の成績は、同時に測定することは不可能である。このように、評価基準が時間的に未来に測定される場合、「予測的妥当性 (predictive validity)」とよばれている。

同時的妥当性でも予測的妥当性でも、どちらの場合も評価方法は、問題となっている尺度と外的基準との相関を検討することにより行われるのが普通である。

◆内容的妥当性

小学校卒業時の算数の計算能力を測る場合、どのような問題を作成すればよいだろうか。もしも、すべての問題が足し算ばかりで、引き算、かけ算、わり算などを含んでいなければ、そのような計算問題は内容に問題があると考えられるだろう。このような場合、その問題は「内容的に妥当ではない」といわれる。このような妥当性のタイプは「内容的妥当性 (content validity)」とよばれている。

一般に、上記の算数の問題例のように、すべての尺度でその内容的妥当性が簡単に評価できるわけではない。高齢者のQOLの内容は何か、患者の不安を構成するものは何かなど、実際には明確に定義できないことも多いだろう。そのような場合には、内容的妥当性は、抽象的な概念の測定に対する、その尺度の内容に関しての努力目標的な意味をもつものと考えられる。

◆構成概念妥当性

そのほかのタイプの妥当性として、「構成概念妥当性 (construct validity)」がある。作成した尺度の測定結果が、ほかの諸概念・構成概念の尺度の測定

結果と、理論的に導かれる仮説に関して、どの程度符合しているかにより評価される。たとえば、QOL尺度を開発した場合、その尺度得点の高い患者では、生活満足度（得点）や社会的適応（得点）が高いはずであり、また情緒不安定（得点）などは低いものと期待されるだろう。

◆尺度の妥当性が疑われた場合

作成した尺度について、その理論的な仮説と実際の測定との間に破綻がある場合、いくつかの理由が考えられるだろう。すなわち、

- (1) 基礎にある理論の誤り
- (2) 理論を検証する方法などの誤り
- (3) 測定の信頼性の低さなど

が考えられよう。

(1) の場合は、尺度構成をする基礎となった理論的な誤りである。たとえば、QOLを「経済的要因と家族関係要因の相互作用で決定できる」という単純なモデルから作成した場合などがそうである。このような場合には、尺度に関係するほかの要因を新たに導入したり、より適切な理論やモデルを考慮する必要がある。

(2) の場合は、比較的多くみられるのではなかろうか。たとえば、高齢者の日常生活活動度（ADL）を測定するための尺度を開発したとしよう。そのADL尺度の妥当性を調べるために、大学生を対象に調査を実施した場合、その調査結果から開発したADL尺度は妥当でないという結果になっても、その尺度を否定する結論にはならないだろう。なぜならば、明らかに調査対象の選択が不適切であり、得られた結果からは、その尺度が本来目的とする事象の測定が可能か否かを評価することはできないからである。

(3) でいう「信頼性の低さ」の問題は、統計理論から、信頼性の低い尺度間の相関係数は、それらの尺度の信頼性を超えないことが明らかとなっている。

したがって、基準関連妥当性などの評価を相関係数で行う場合、信頼性の低い尺度を用いると、見た目の相関があまり大きくならないことがある。

実際には、(1)～(3)のどれが尺度の妥当性を脅かしている原因かを特定することは、困難な場合が少なくない。そのような場合、尺度構成に使用した項目を用いて因子分析（多変量解析の方法の一つで、変数間の相関係数を用いて、多数の変数間に共通する因子が何かを検討するための分析方法）などを行うことで、項目間の構造を検討したり、複数の尺度間の解析を行うのも、妥当性を評価するうえでは有用なものとなることもある。

尺度構成での信頼性について

尺度構成における「信頼性 (reliability)」とは、同一項目に関して繰り返し測定をした場合の一致性を示すものである。したがって、2つの反復する測定の結果が一致적であれば、その尺度の信頼性は高いことになる。以下に、簡単に信頼性を測るための統計理論を解説しよう。

◆クロンバッハの α 信頼性係数

ある事象の測定において、真の得点を t 、測定得点を x 、測定誤差を e とすると、 $x = t + e$ 、と書くことができる。このとき、信頼性 ρ は、真の得点 t と測定得点 x の分散の比、 $\rho = [t \text{ の分散}] / [x \text{ の分散}]$ 、によって定義される。

実際には、真の得点を直接測定することは不可能である。そのため、2つの同等な尺度の同時的な測定（平行測定）や、調査の反復測定（再テスト法）により、2つの繰り返し測定された尺度の相関係数を求めることで、信頼性の推定を行うことができる。

1回の測定で信頼性を推定するには、以下のような方法を用いる。なお、ここでは、複数の項目への回答からの合成得点（「スコア」とよぶこともある）に

ついて、信頼性係数を求める方法について解説しよう。

今、 p 個の変数 $X_1, X_2, \dots, X_p$ の合計点（尺度得点、スコア）を Y 、とする。各変数 X_i ($i=1, 2, \dots, p$) と合計点 Y の分散を V_i, V_Y としよう。一般によく用いられている「クロンバッハ（「クロンバック」ともよばれる）の α 信頼性係数（Cronbach's α ）」は、

$$\alpha = \left[\frac{p}{p-1} \right] \times \left[1 - \frac{\sum_i V_i}{V_Y} \right]$$

によって求められる。なお、 $\sum_i$ は p 個の変数すべてについての和を求めることを意味している。

なお、クロンバッハの α 信頼性係数を求める場合、尺度得点の計算に用いる項目間の相関係数は、すべて正である必要がある。ほかの項目と負の相関である項目は、尺度得点の計算に用いないか、得点の付け方を逆にしなければならない。

◆その他の信頼性係数

尺度の信頼性係数は、 α 信頼性係数のほかにもいくつか提唱されている。比較的によく用いられるほかの信頼性係数に θ 信頼性係数がある。

多くの変数から合成変数を求める多変量解析の方法として、主成分分析がよく知られている。主成分分析の第1主成分は、分析に用いた変数の重みづけの合成得点であり、最適な尺度構成を行った場合に対応している。通常の尺度構成では、各変数にかかる重みは1に設定されているのだが、主成分分析では変数間の相関構造を考慮して、最適な重みを求めているのである。

変数間の相関係数行列を用いた主成分分析において、第1主成分の固有値を λ とした場合、 θ 信頼性係数は、

$$\theta = \left[\frac{p}{p-1} \right] \times \left[1 - \frac{1}{\lambda} \right]$$

によって求められる。

θ 信頼性係数は、用いられる変数の最適な重みづ

けをした場合の信頼性なので、 α 信頼性係数とは、 $\alpha \leq \theta$ 、の関係があることが知られている。

◆尺度構成での変数選択について

尺度得点の計算に用いる各項目が、その尺度と関係があることが尺度構成の前提となっている。しかし、当初の項目の設定の考えに反して、実際のデータでは、ある項目は作成したい尺度とそれほど関係していない場合もありうる。そのような状況を検討し、尺度構成に必要な項目を選択するための方法がいくつかある。

最も簡単な方法は、すべての項目を用いた場合の α 信頼性係数と、各項目についてその項目を除いた場合の α 信頼性係数を計算し、その違いを比較すればよい。すなわち、その項目を除くと α 信頼性係数の減少幅が大きな項目は、その尺度に重要な役割をもつと考えてよい。逆に、その項目を除いたほうが α の値が大きくなるようならば、その項目は用いないほうがよいことになる。

また、前述のように、信頼性係数を求めるのに、主成分分析を利用することもできる。尺度の項目間に相互に高い相関があり、その和を計算できるのであれば、すべての項目の第1主成分の負荷量（主成分との相関係数）が正に大きくなるはずである。もしも第1主成分が負の項目がある場合には、その項目の得点の付け方は逆順にして尺度得点の計算に用いなければならない。また、負荷量が0に近いような項目はその尺度の計算に用いる意味がないことを示している。

②は①に示した共感性尺度について、約32,500人の成人男女について調査をした結果である。6項目での α 信頼性係数は0.662、 θ 信頼性係数は0.693であった。この α と θ の値が近いほど、用いた項目での尺度化が最適に行われていることを示すものと考えられる。

②の「 α 係数」は、当該項目を一つ除いた場合の α

項 目	α 係数	負荷量
1) 他人の気持ちが人一倍よくわかる	0.590	0.719
2) たとえ自分は損をしても友人の苦境は見過ごせない	0.600	0.676
3) 他人の苦しみがよくわかる	0.580	0.745
4) 人のために尽くすことは好きだ	0.598	0.681
5) 人一倍感受性が豊かである	0.648	0.501
6) 捨て犬、捨て猫を見捨てるのはつらい	0.698	0.355

② 共感性尺度項目の検討

項 目	α 係数	負荷量
1) 他人の気持ちが人一倍よくわかる	0.624	0.733
2) たとえ自分は損をしても友人の苦境は見過ごせない	0.648	0.678
3) 他人の苦しみがよくわかる	0.615	0.756
4) 人のために尽くすことは好きだ	0.646	0.684
5) 人一倍感受性が豊かである	0.707	0.507

③ 項目6)を除いた場合

信頼性係数を示している。この値が大きいほどその項目の尺度への寄与が小さいことを示しており、全項目を用いた α 信頼性係数(0.662)より大きければ、その項目は尺度構成に使用しないほうがよいことになる。②の「負荷量」は、第1主成分負荷量を示しており、この値が小さいほど尺度構成への寄与が小さいことを示している。したがって、②の「6) 捨て犬、捨て猫を見捨てるのはつらい」は、 α が0.698と大きく、負荷量は0.355と小さいので、共感性尺度から除いたほうがよい項目ということがわかる。

③は、項目6)を除いた場合の結果を示したものであり、5項目の場合の α 信頼性係数は0.698、 θ 信頼性係数は0.705となる。②の場合と同様に、③の結果から、「5) 人一倍感受性が豊かである」という項目

項 目	α 係数	負荷量
1) 他人の気持ちが人一倍よくわかる	0.645	0.732
2) たとえ自分は損をしても友人の苦境は見過ごせない	0.661	0.699
3) 他人の苦しみがよくわかる	0.610	0.778
4) 人のために尽くすことは好きだ	0.658	0.708

④ さらに項目5)を除いた場合

を除いたほうが α 信頼性係数が大きくなることがわかる。

④は、さらに一つ項目を減らして4項目のみで尺度化を行った場合についての結果である。全体では、 α 信頼性係数は0.707、 θ 信頼性係数は0.708とほぼ等しい結果になった。④の結果では、これ以上どの項目を除いても α 信頼性係数の値は大きくならないので、変数の削除はこれで終了である。このときの、各項目の主成分負荷量は0.699～0.778とかなり大きな数値になっており、4項目ともにすべて第1主成分との相関が大きいことがわかる。

尺度構成に関するそのほかの話題

尺度構成は、いくつかの項目について対象者に回答してもらい、その総合得点である事象を測定するものである。したがって、できるだけ少ない項目で測定ができれば、それに越したことはない。ただし、通常は項目数が大きければ α 信頼性係数などがある程度大きくすることはできるが、極端に項目数が少なくなると信頼性係数も小さくなるのが普通である。このため、尺度構成の作業開始時には、上記の例のように、ある程度多くの項目を用いて検討を行い、信頼性のある程度維持したまま(0.8以上の信頼性があれば望ましい)で項目数を減らすという作業が必要である。

また、測定したい事象に関して下位の概念が存在する場合、調査データから仮定した概念がうまく抽出できるのかを、因子分析で検討することもよく行われる。この点についての解説は、誌面の都合で省略するが、現実の尺度構成ではかなり重要な分析となることが多いだろう。

ところで、構成した尺度を用いて、対象者に関する解析を行うのが普通であるが、その場合、尺度得点は量的データとして解析されるのが普通である。すなわち、尺度構成とは、順序尺度で構成された多数の項目から、間隔尺度の量的データとなるような「ものさし」を作る方法といえるものである。したがって、尺度得点の平均値や分散を求めて、量的データに基づく統計学的検定を行うのに特に問題はない。しかし、統計学的検定の基礎として、尺度得点の母

集団での分布が正規分布に従うか不明であることから、ノンパラメトリック検定を用いる例も見受けられるが、現実には両者の検定結果などに大差がないことも多い。

尺度構成では、作成している尺度の信頼性の検討を行う場合、多変量解析の手法を用いるので、理解するのがそれほど容易ではないかもしれない。また計算の方法もそれほど簡単ではないので、解析や項目の検討のために適当な解析ソフトが必要である。今回の計算はすべて、統計学ソフトHALWINの一部である尺度構成に関するプログラムを用いて行った。興味ある読者は、筆者のホームページ (<http://www.clg.niigata-u.ac.jp/~takagi/>) を参照していただければよいだろう。

参考文献

- 水野欣司、野嶋栄一郎訳：テストの信頼性と妥当性。朝倉書店；1983。（Carmines EG, Zeller RA: Reliability and Validity Assessment. Beverly Hills: SAGE Pub; 1979.）
高木廣文：HALWINによるデータ解析。現代数学社；1998。
高木廣文、柳井晴夫：生活習慣の信頼性と因子構造の検討。統計数理 1998; 46(1): 39-64。
高木廣文：測れないものを測る方法—尺度化について—。超音波検査技術 2000; 25(4): 275-280。