

多変量解析を理解するために(6)

尺度構成法の基礎

高木 廣文 | Takagi Hirofumi

東邦大学医学部看護学科国際保健看護学研究室教授

■1950年生まれ。74年東京大学医学部保健学科卒。79年同大学院博士課程修了。同年米国国立衛生院奨励研究員として米国国立環境保健学研究所勤務。81年聖路加看護大学講師、82年同大助教授。89年文部省統計数理研究所助教授。99年新潟大学医療技術短期大学部看護学科教授、同年同大医学部保健学科教授。2006年4月より現職。著書に『ナースのための統計学 ― データのとり方生かし方』(医学書院)、『多変量解析ハンドブック』(現代数学会)、『健康科学とコンピュータ』(共立出版)など。

1. 「こころ」の中を測る方法

身長や体重を測定するように、「知能」を知能指数で本当に測れるのだろうか。手術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、いざ測定しようとなると極めて難しい事象が、現実には沢山ある。そのような事象を測定するための専用の「物差し」を作成する方法は、「尺度構成法 scaling」として知られている。

ある事象を測るための専用の「物差し」は「尺度 scale」と呼ばれる。新しい尺度を作るには、その概念に関係すると考えられる項目を多数作成し、それらの項目に対する回答に適切な得点を与え、その合計得点でその事象を測定すればよい。尺度構成とは、順序尺度から成る多数の項目を用いて、「間隔尺度を構成する方法」と解釈することができる。

表1は、共感性尺度の6項目を示したものである。「3. はい」「2. どちらでもない」「1. いいえ」から1つだけ回答を選んでもらい、

表1 共感性尺度項目

- | |
|----------------------------|
| 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みすごせない |
| 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) すて犬、すて猫を見捨てるのはつらい |

各回答に対して、3、2、1の得点を与えて、その合計点で共感性を測定することにする。尺度構成で最も重要な点は、作成した尺度が目的とする概念を正しく適切に測定しているのかということである。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。

2. 尺度の妥当性について

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、以下のように複数のタイプに細分して考えることができる。

(1) 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。ひとつの考え方として、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、質問紙を用いて塩分の摂取量を推定した場合、実際の塩分摂取量との相関が高ければ、質問紙を用いた塩分摂取量推定法は妥当であると考えてよいだろう。このように、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」と呼ばれている。

基準関連妥当性は、尺度と外的基準の測定時期によって、さらに2つに分類される。質問紙法による推定塩分摂取量と栄養調査による塩分摂取量の妥当性を検討する場合、2つの測定は同時に行われるので、「同時的妥当性 concurrent validity」と呼ばれる。一方、大学入学試験と入学後の成績との相関を調べる場合、評価基準である入学後の成績は時間的に未来にあるため、「予測的妥当性 predictive validity」と呼ばれる。

(2) 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、すべての問題が足し算ばかりで、引き算、掛け算、割り算を含んでいなければ、その計算問題は内容に問題があり、「内容的に妥当ではない」と言われる。このような妥当性のタイプは「内容的妥当性 content validity」と呼ばれる。

内容的妥当性の確認のために、その分野の専門家に尺度項目をチェックしてもらうのもひとつの方法である。下位尺度がある場合には、複数の専門家に全項目を各下位尺度に分類してもらい、各項目ごとの専門家による分類の一致度を検討するという方法もある。一致度が低ければ、尺度項目としては適切なものとは言えない。

一般に、すべての尺度で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 (Quality of Life) の内容は何か、患者の不安を構成するものは何かなど、明確に定義できないことも多い。一般に、内容的妥当性は抽象的な概念の測定に対する努力目標的な意味を持つ。

(3) 構成概念妥当性 construct validity

作成した尺度の測定結果は、他の諸概念・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。例えば、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。

(4) 尺度の妥当性が疑われた場合

主な理由として、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、などが考えられる。

①は、尺度構成の基礎となった理論上の誤りで、例えばQOLは「経済的要因と家族形態で決まる」というような単純なモデルを基にした場合がそうである。このような場合、より適切な理論やモデルを考慮し、尺度に関係

する他の要因を新たに導入する必要がある。

②は、例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場合、ある事業所での従業者を対象に調査した結果から、そのADL尺度は妥当でないという結果が示されても、その尺度を否定する結論にはならない。その尺度が目的とする対象集団ではないために、その尺度の評価はできないからである。

③の「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」ことが知られているからである。基準関連妥当性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、得られた相関があまり大きくなることがある。

実際には、尺度の妥当性を脅かしている原因を特定するのは極めて難しい。尺度構成に使用した項目を用いて因子分析を行うことで、項目間の構造を検討したり、複数の尺度間の解析を行うことも、妥当性を評価する上では有用なものとなる。

3. 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。従って、2つの反復する測定の結果が一致的であれば、その尺度の信頼性は高いことになる。

(1) クロンバックの 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、

$$x = t + e$$

と書ける。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、

$$\rho = \frac{V(t)}{V(x)}$$

によって定義される。ここで、 $V(t)$ 、 $V(x)$ は真の得点 t と測定得点 x の分散をそれぞれ示す。

実際は、真の得点 t を直接測定することは不可能なので、2つの同等な尺度の同時的な測定(平行測定)や、反復測定(再テスト法 test-retest method)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、複数の項目への回答から求められた合成得点(スコア)の信頼性係数を求める方法は、以下のようなものである。

p 個の変数 $x_1, x_2, \dots, x_p$ の合成得点を y とし、各変数 x_i ($i=1, 2, \dots, p$)および合成得点 y の分散を $V(x_i)$ および $V(y)$ とする。一般によく用いられている「クロンバック(クロンバッハとも呼ばれる)の α 信頼性係数Cronbach's α 」は、

$$\alpha = \frac{p}{p-1} \times \left[1 - \frac{\sum_{i=1}^p V(x_i)}{V(y)} \right]$$

によって求められる。なお、クロンバックの α 信頼性係数を求める場合、尺度項目間の相関係数は、すべて正でなければならない。他の項目と負の相関となる項目は、尺度得点の計算に用いないか、得点の付け方を逆順にする。

(2) θ 信頼性係数

多くの変数から合成変数を求める方法に主成分分析がある。主成分分析の第1主成分は、すべての変数の重み付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、各変数に掛かる重みは1に設定されているが、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。

変数間の相関係数行列を用いた主成分分析の第1主成分の固有値を λ とすると、 θ 信頼性係数は、

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

によって求められる。

θ 信頼性係数と α 信頼性係数との間には、 $\alpha \leq \theta$ 、の関係があることが知られている。

(3) 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は、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が、ある項目はその尺度とそれほど関係していない場合もあり得る。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幾つかある。

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、その差を比較するという方法である。その項目を除いた場合、 α 信頼性係数の減少幅が大きな項目は、重要な役割を持つ。逆に、その項目を除いた場合、 α の値が大きくなるのなら、その項目は用いない方がよい。

また、主成分分析の結果を利用することもできる。尺度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。第1主成分に負の項目があれば、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る。また、負荷量が0に近い項目は、その尺度の計算に用いる意味がないことを示している。

表2は、表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。 α 信頼性係数と θ 信頼性係数の値が近いほど、尺度化が最適に行われていることを示すものである。

表2中の「 α 係数」は、その項目を1つ除いた場合の α 信頼性係数を示している。この値が大きい項目ほど、尺度への寄与が小さいことを示し、全項目での 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方がよい。同様に表2中の「負荷量」は、第1主成分負荷量を示し、この値が小さいほど尺度構成への寄与が小さい。従って、表2中の α と θ の値は、大小関係が逆になっている。

表2の結果から、「6) すて犬、す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方がよい項目ということになる。

表2 共感性尺度項目の α 係数と主成分負荷量 ($\alpha=0.662$, $\theta=0.693$)

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みすごせない	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) すて犬、すて猫を見捨てるのはつらい	0.698	0.355

表3は、項目6)を除いた場合の結果である。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。表2の場合と同様に、表3の結果からは、「5) 人一倍感受性が豊かである」という項目を除いた方が α 信頼性係数が大きくなる。

表4は、さらに1つ項目を減らして4項目のみで尺度化を行った場合の結果である。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。表4で分かるように、各項目の α 信頼性係数は、全項目での α 信頼性係数(0.707)より小さい。従って、どの項目を除いても α 信頼性係数の値は、これ以上大きくならないので、変数の削除はこれで終了である。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、すべての項目が第1主成分との相関が大きいことが分かる。

4. おわりに

尺度構成は、心理学はもとより医学、看護学でも広く使用されている。人間の心の中は原理的に解明が不可能であるが、特定概念に適切な構造を仮定することで、量的な把握が可能となり、現象解明に役立つ場合も多い。

多変量解析の各方法は、現実世界の複雑な現象の解明のために考え出された方法であると言えるだろう。この連載も今回限りとなったが、少しはデータ解析を志す人たちの役に立ってくれることを願っている。

表3 1項目除いた場合 ($\alpha=0.698$, $\theta=0.705$)

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みすごせない	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

表4 さらに1項目除いた場合 ($\alpha=0.707$, $\theta=0.708$)

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みすごせない	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

* 参考文献

- [1] 高木廣文・林邦彦(2006): エビデンスのための看護研究の読み方・進め方: 中山書店.
- [2] 水野欣司・野嶋栄一郎訳(1983): テストの信頼性と妥当性: 朝倉書店 (Carmine EG, Zeller RA: Reliability and Validity Assessment. Beverly Hills: SAGE Pub; 1979).
- [3] 高木廣文・柳井晴夫(1998): 生活習慣の信頼性と因子構造の検討: 統計数理, vol.46(1): pp.39-64.
- [4] 高木廣文(2000): 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, 25(4), pp.275-280.
- [5] 高木廣文(2007): 多変量解析を理解するために (4)主成分分析の基礎: エストレーラ.
- [6] 高木廣文(2007): 多変量解析を理解するために (5)因子分析の基礎: エストレーラ.
- [7] 柳井晴夫・高木廣文編著(1986): 多変量解析ハンドブック: 現代数学社.
- [8] 高木廣文・柳井晴夫編著(1995): HALBAUによる多変量解析の実践: 現代数学社.
- [9] 高木廣文(2006): HALBAU7によるデータ解析: シミック(株).