

標本抽出法が心理検査の標準得点に及ぼす影響

村上 宣寛¹・村上 千恵子²

The influence of sampling methods on standard scores of psychological tests

Yoshihiro MURAKAMI, Chieko MURAKAMI

E-mail: completewalker@gmail.com

chiekomurakami.jp@gmail.com

概要

本論文の目的は、日本で行われてきた心理検査の標準化に用いられた標本抽出法の批判的検討である。大部分の心理検査は日本人全体から無作為に抽出された標本で標準化されていない。無作為抽出されていない標本では、母集団の統計的推測が不可能になる。標本を大学生や健常者に限定した場合は、切断正規分布になりがちで、素点の分散が小さくなる。Z得点法ではその逆数を利用するので、標準得点の分散が大きくなる。そのような心理検査を日本人一般に適用すると、標準得点の分散が大きいため、異常な低得点や高得点が頻出する。本論文では、MMPI-1と新日本版MMPIでその影響を具体的に示した。アメリカのMMPIの凋落を招いた要因には、古いミネソタ基準の標本がアメリカ人という母集団を正しく代表していなかったこと、質問項目に攻撃的・差別的表現が含まれていたこと、精神医学界でDSMという診断基準が普及したので、大規模質問紙の需要がなくなったことなどがある。

キーワード：標本抽出法, Z得点, 正規化T得点, MMPI

Keywords : sampling, Z-score, normalized T-score, MMPI

はじめに

推測統計学は、母集団の統計的性質を標本から推測する学問である。母集団の統計的性質は直接的に得られないため、母集団の性質をよく代表し、その規模を縮小した標本が必要になる。数学的にはそれは無作為抽出法による標本である。

社会調査では無作為抽出法に準じた方法が考案され、使用されてきた。一方、心理学では、標本抽出法に配慮した研究は少ない。大学生を標本とした研究は、大学生に当てはまる程度である。

精神神経科で用いられる心理検査は、正常／異常の弁別目的で作成されている。例えば、BDI-IIは13／14点を正常／抑うつを弁別する得点としている¹⁾。GHQも同じで、神経症者／健常者を弁別する得点を16／17点としている²⁾。言い換えれば、これらの心理検査はその得点の高低で解釈するようには設計されていない。また、投影法や作業検査法

は標準化の概念がない。ここでは、このような心理検査は除外する。

心理検査の得点は伝統的に平均50、標準偏差10で標準化される。いわゆる偏差値と同じである。標本が母集団から無作為抽出法に準じた方法で抽出されていれば、標準得点が60点であれば、正規分布表から上位確率0.15866の高得点、70点であれば上位確率0.02275、80点であれば上位確率0.00135、90点であれば上位確率0.00000になる。すなわち、心理検査で標準得点の70点は100名に2名程度の高得点、80点は1000名に1名程度の非常な高得点、90点は理論確率ゼロの超高得点である。

標本抽出法

母集団から一部を選び出し、調査する方法を標本調査、取り出された一部を標本(サンプル)、取り出す手続きを標本抽出(サンプリング)と呼ぶ。標本調査ではデータを分析し、その結果から母集団の性質を推測する。したがって、標本は母集団の忠実な縮図である必要がある。

¹ 富山大学名誉教授

² 元信州大学教授

母集団が限定された少人数の母集団であれば、無作為抽出法が可能である。ただ、母集団が数万人の社会調査や、日本人全体を母集団とする心理検査の標準化などでは、無作為抽出法は技術的に実行不可能である。

以下に主な標本抽出法を説明する。

- **無作為抽出法**... ランダム・サンプリングとも呼ぶ。母集団から各標本をまったく同じ確率で選ぶ方法である。適当な、でたらめな方法では無作為抽出にならないのが普通である。具体的には乱数表やくじを利用して標本抽出を行う。統計技法の適用には理想的な方法である。ただし、単純で簡単な方法に思えるが、母集団が日本人全体というように大規模集団になると、無作為抽出は難しい。日本人全体の戸籍を取り寄せて乱数表が割り当てられるだろうか。質問紙法では行われた例がない。
- **系統抽出法**... 母集団の規模が大きい場合、単純な無作為抽出法を適用すると、作業量が増えて実行困難になることがある。そこで、母集団に1 から番号を付け、等間隔に抽出する方法がある。例えば、10000 名から100 名を抽出する場合、 $100/10000 = 1/100$ の確率で調査対象者を抽出すればよい。最初に乱数表を見て、スタート番号（例えば56）を決め、100 番目ごとに（156、256... と）抽出すると、ちょうど100 名の調査対象者が選ばれる。数千名規模の社会学的調査でよく利用される。
- **層化抽出法**... 職業、収入、教育歴、世代など、調査内容に影響しそうな統計カテゴリーで、母集団をいくつかの階層に分け、各階層から調査対象者を抽出する方法である。一般的に、階層の人口比に比例させる比例割り当て法が用いられる。階層の内部では、統計値がなるべく等しくなるように、階層の間では、統計値の相違がなるべく大きくなるように階層を決める必要がある。無作為抽出法や系統抽出法と比べると、少ない標本でも測定精度を上げることができる。社会学的調査でよく利用される。

- **多段抽出法**... 全国調査などの場合、無作為抽出法や系統抽出法を適用すると、調査員が広い地域を移動しなければならない。そこで、まず大都市、小都市、市町村などを抽出し、次に住民票から調査対象者を抽出するという方法をとる。手続的には等確率抽出法と確率比例抽出法に分かれる。どちらの方法でも個人が調査対象者に選ばれる確率は同じである。等確率抽出法は、各都市を無作為に（等確率で）抽出し、その後、調査対象者を人口比に合わせて抽出する方法である。確率比例抽出法は、各都市を人口比に応じた確率で抽出し、その後、調査対象者を無作為に同数ずつ抽出する方法である。なお、全国調査などの場合は、国勢調査と比較するため、調査対象者の教育状況、職業などについての質問項目を含めておく。そうすれば、それらの割合が国勢調査と同じであるかが検証できる。

系統抽出法、層化抽出法、多段抽出法は無作為抽出法を大集団で可能にする方法であり、これらの方法で標本を抽出すると、それは母集団の縮図になり、統計的推定が可能になる。逆に、無作為化が保証されない他の抽出法では、母集団の統計的性質の推定ができない。

なぜ無作為抽出法が必要なのか

無作為抽出では各標本が母集団から選ばれる確率が同じである。無作為抽出によって選ばれた標本平均 \bar{x} の分布について、次の定理がある³⁾。

定理 1 平均 μ 、標準偏差 σ の正規母集団から無作為抽出された大きさ n の標本平均 \bar{x} は、平均 μ 、標準偏差 σ / \sqrt{n} の正規分布に従う。

この定理は純粹に数学的に導かれたもので、標本平均から母平均を推定する精度が、標本の大きさ $1/\sqrt{n}$ に依存することを示している。

定理 2 平均 μ 、標準偏差 σ のある母集団から無作為抽出された大きさ n の標本平均 \bar{x} は、 n が無限に大きくなると、平均 μ 、標準偏差

σ / \sqrt{n} の正規分布に近づく。

この定理は中心極限定理 (central limit theorem) として有名である。この定理を使えば、正規分布からかなりずれた分布でも、標本から母集団の平均や標準偏差が推定できる。

無作為抽出された場合にのみ、標本から母集団の統計的性質が推論可能である。逆に、無作為抽出法でないと、標本を分析しても母集団の統計的性質は推定不能である。

ホーエル (1981, P.123) ⁴⁾ は「一般的に、無作為抽出以外の他の抽出法の結果に関しては正しい意味での確率的表現を与えることはできない。統計家が標本は無作為に選ばなければならないと主張するのはこの理由からである」という。

また、盛山 (2004, Pp.117-118) ⁵⁾ も「サンプリングは一般に無作為抽出 (random sampling) でなければならない。... 無作為抽出をフォーマルに定義すれば、それは母集団の中のすべての個体にとって、標本として抽出される確率が等しいものになっているような抽出法である。... 抽出が等確率であることによって、中心極限定理などの確率理論を利用することができ、それによって母集団特性をリーズナブルに推論できるのである。そのため、無作為抽出は確率抽出 (probability sampling) とよばれる。」盛山は等確率で抽出する方法として、単純無作為抽出、多段抽出法、層別抽出法、系統抽出法を挙げ、それ以外の抽出法は無作為抽出ではないという。

正規分布とは

一般的な測定値は、度数分布表の中央の頻度が高く、周辺に隔たるにつれて標本数は減る場合が多い。標本数が大きい場合、特にこのような傾向が見られる。このような釣り鐘型の分布を正規分布 (normal distribution) と呼ぶ。

正規分布曲線の形は平均値と標準偏差の二つの値によって決定される。平均値 μ は中央の値であり、標準偏差は変曲点という曲線の向きが変わる場所を示す。また、 $\mu \pm 1\sigma$ の間に 68.3% が、 $\mu \pm 2\sigma$ に 95.4% が属する。つまり、標本の値を知るだけで全体のどの位置にあるかが分かる。

Z 得点による標準化

平均値や標準偏差の測定単位が異なると、全体での位置はわからない。それで、分布の単位をそろえるために標準化を行う。

z 得点は、

$$z = \frac{x - \bar{x}}{\sigma}$$

で定義される。 z は常に平均 0、標準偏差 1 となる。なお、有名な偏差値は Z 得点の俗称で、 $10z + 50$ で変換したものである。これが心理検査で多用される標準得点である。

z の定義式を見て分かるように、 z は標本平均を標本の標準偏差 (分散) で除した形である。無作為抽出法などに基づかないと、標本の標準偏差 (分散) が小さくなり、標準得点の分散が大きくなる。大学生で行った標準化例などを後述する。

正規化 T 得点

得点分布が正規分布から離れると、標準得点と全体での位置が対応しなくなる。つまり、標準得点で 70 点ならパーセンタイルで 98、つまり、100 名に 2 名程度の出現頻度なのに、10 名に 1 名とか、頻度が大きく変わってしまうことがある。すると、70 点は異常値とは見なせないの、解釈も変更する必要がある。このような場合に得点分布を正規分布に変更して標準得点を求める。池田 (1994) ⁶⁾ のパーセンタイル順位から T 得点に換算する表を使うと簡単である。ここではクライン (1986) ⁷⁾ に準じて手順を示す。

1. 度数分布表から累積度数を求める。累積度数は特定の素点以下の度数を合計したものである。
2. 累積度数から各得点幅の中間点を求める。これは各得点における度数の半分を各得点の累積度数に加える。例えば、素点 5 の累積度数が 10 で、素点 5 の被検者が 4 名の場合、素点 6 の中間点は 12 である。
3. 累積度数の中間点を総頻度で割り、累積率を求める。

4. 累積率は正規分布表の面積に該当するので、正規分布表から z を読みとる。 z 以下の面積を表示している表を使うと便利である。0～ z の面積しか表示されていない場合は、累積比率から 0.5 をひいた値で見る。
5. 求めた z を線形変換する。性格検査の場合は $T = 10z + 50$ である。

知能の分布のように正規分布が理論的に仮定される場合や、正規分布からの逸脱が大きく解釈に支障が出る場合などは T 得点化した方がよいが、もとの分布を歪めてしまう。

Z 得点と T 得点の違いを簡単にまとめておこう。

- 得点分布がほぼ正規分布している場合、Z 得点と T 得点はほぼ同じである。このため、Z 得点による標準化にも関わらず、アメリカの MMPI のように T 得点と呼ぶことがある。
- 得点分布が正規分布から外れている場合、Z 得点と T 得点は大きく食い違う。Z 得点を使うと、高得点者や低得点者が異様に少なくなったり、多くなったりして、解釈に支障を来す。T 得点ではこのような弊害はない。
- 得点分布に正規分布が仮定できない場合、分布の歪み自体が何らかの心理的事実を物語っているかもしれない。このような場合、T 得点化を行うと分布を正規化してしまい、その心理的事実は覆い隠されてしまう。T 得点化が常に望ましいわけではない。

大学生標本の特徴

大学生は一般的に学力の度合いに応じて選抜された集団である。正規分布していた集団の一部を切り取るので、この場合は、切断正規分布と呼ぶ。切断の場所によって効果が異なるが、標準偏差（分散）は選抜前の集団より必ず小さくなる⁸⁾。

心理学分野では選抜集団で相関係数を求めると小さな値になるので、これを切断効果と呼んでいる。大学入試の得点と大学での成績の相関を求める研究

がよく行われるが、決まって相関係数は小さな値になる。これは大学入学者だけで相関関係を調査するからである。真の相関を求めるには不合格者の大学での成績が必要であるが、これは存在しない。

大多数の大学生の学力偏差値は 50 以上である。精神障害にかかっている確率はかなり小さい。例えば、世界精神保健日本調査セカンドの大規模疫学調査⁹⁾では、住民基本台帳をもとに多段階無作為抽出を行って精神疾患の有病率等を調べた。2013～2015 年度までに関東地方調査、東日本調査、西日本調査、関東地方補充調査で、2450 名に面接調査と自己記入式調査票による調査を行った。生涯有病率の設定に回答した被験者 2288 名のうち、精神病症状を経験した人は、中学・高校卒で 4.02%、専門・短大卒で 4.15%、大学・大学院卒で 2.28%であった。精神疾患の有病率は大学・大学院卒群で小さい。

大学生は均質なので、大学生で算出した標準偏差は小さめで、日本人全体から無作為抽出した標準偏差は大きめになる。Z 得点は $Z = 10 \times \frac{x - \bar{x}}{\sigma} + 50$ で計算されるので、この σ が小さくなると、標本の値と標本平均の差を標準偏差で除した $(x - \bar{x})/\sigma$ が大きくなる。

大学生は日本人全体から等確率で抽出された標本ではない。大学生で算出した標準偏差を、それを適用してはいけない日本人全体という集団で用いると、標準得点の分散が大きくなり、低得点や高得点が頻出する。推測統計の誤用である。大学生で標準化された心理検査の適用範囲はせいぜい大学生である。

大学生標本による標準得点の歪み

当初、MMPI-1 は大学生標本に基づいて、標準化されていた。大学生標本のデータは、男性 118 名（平均年齢 20.77、標準偏差 1.43）、女性 201 名（平均年齢 20.16、標準偏差 0.93）であった¹⁰⁾。図 1 は、事例に紹介した精神分裂病の分類不能型（DSMIII-R による医師の診断）の MMPI-1 の基本プロフィールである。

その後、MMPI-1 は住民票からの多段階無作為抽出法を採用した。四国、沖縄を除いて全国を 7 ブロックに分け、15 都道府県、30 区市町村、80 小学校区を抽出し、男女 4700 名を無作為抽出し、郵送法で回答を求めた。有効回答数は男性 494 名、女性 684

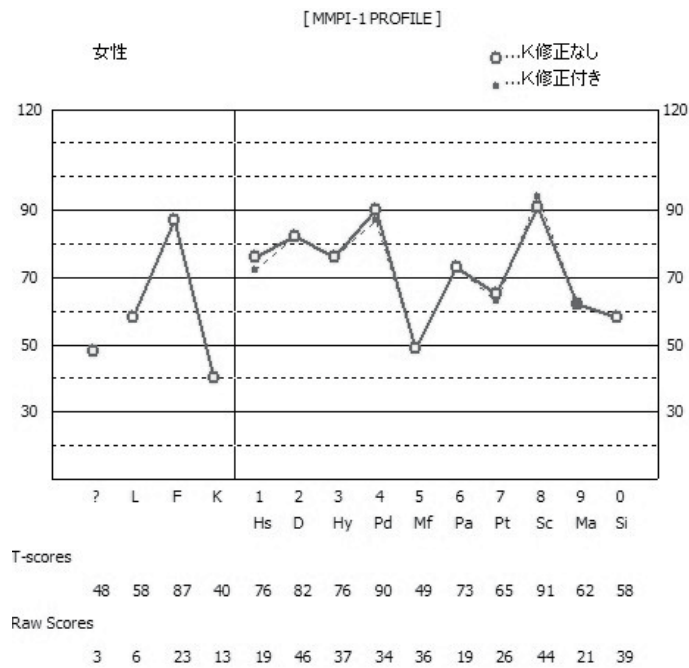


図 1: 大学生基準 (村上・村上, 1992) に基づく精神分裂病分類不能型 (DSM-III-R による診断) の MMPI-1 プロファイル。

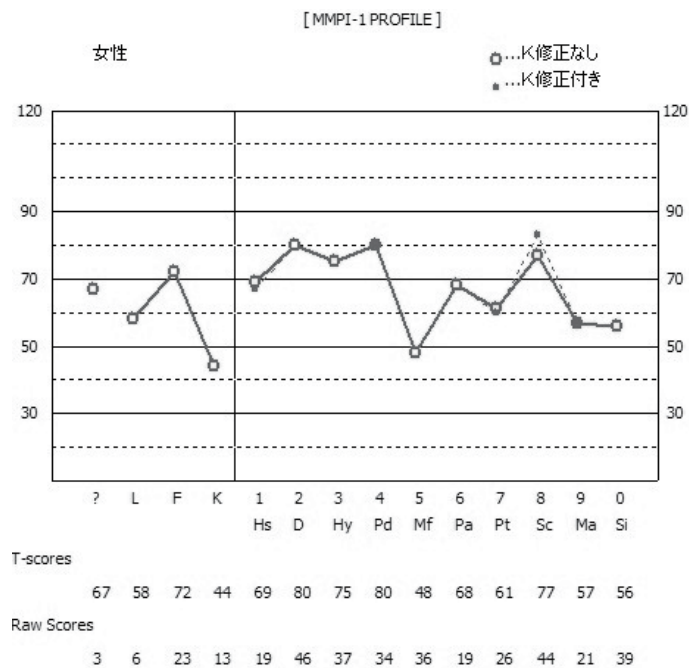


図 2: 多段階無作為抽出法による基準 (村上・村上, 1996) から作成した精神分裂病分類不能型 (DSM-III-R による診断) の MMPI-1 プロファイル

名で、青年期 15～22 歳（男性 122 名、女性 187 名）、成人前期 23～39 歳（男性 107 名、女性 218 名）、成人中期 40～59 歳（男性 126 名、女性 168 名）、成人後期 60 歳以上（男性 139 名、女性 111 名）で標準化を行った。教育歴や職業などの構成は、国勢調査の結果とほぼ一致していた。また、不応答の平均は 0.2～0.7 で、アメリカと同様に 0 に近かった¹¹⁾。世代別に標準化した質問紙法は世界でも例がない。

図 2 は、同じ事例の MMP1-1 の基本プロフィールである。

なお、MMPI-1 自動診断システムのソフトウェアには隠しコマンドがあり、年齢欄に 0 を入れると大学生標本、年齢を正しく入力すると世代別標準化に基づくプロフィールを作成する。図 1 と図 2 はこのコマンドで作成した。

二つのプロフィールを比較すると、大学生を標準化集団にした場合は、高得点が強調され、標準得点で 90 点を超える尺度が二つある。正規分布表から 90 点の理論確率を求めると¹²⁾、0.000032 になる。これは理論確率がほぼゼロであり、あり得ない値である。臨床尺度間に相関はあるが、理論確率がほぼゼロの尺度が二つもある。尺度得点が真であれば、このような患者に遭遇する確率はゼロである。したがって、この尺度得点は偽であり、大学生を標準化集団にしたために、人為的に高得点が得られたのである。

標本抽出法不明の質問紙

多くの質問紙法では、どのように標本を抽出したのか、マニュアルに明記されていない。言い換えると、単純無作為抽出、多段抽出法、層別抽出法、系統抽出法のいずれかであれば、その手続きの記載が不可欠である。というのは、手順が複雑で、さまざまな方法があるからである。マニュアルにそのような抽出手続きが記載されていない場合は、無作為抽出ではなく、機会サンプリング¹³⁾の一種である。

サール (2005, Pp.67-68) によると、機会サンプリングは「研究者が自分が所属している社会集団や文化集団の人に依頼」していく方法である。「無作為サンプリングと機会サンプリングは同じではない。」機会サンプリングでは最初に出会った人たちに被検者を頼むので、母集団全員の抽出確率は異なる。す

なわち、確率抽出（無作為抽出法）ではないので、母集団を日本人全体とすることはできない。

盛山 (2004, P.118) は「無作為抽出でないとするればどうなるか。... いくら人数が多くても同じだ。かりに母集団は日本の有権者全体だと明確になっているとしよう。... 対象者として、繁華街から、ラジオのリリスナーから、... さまざまな手段を駆使して 10 万人ほどの大量のデータを集めたとしよう。この標本データで、内閣支持率が 52.1% だったとする。このデータの実事から、日本の有権者全体の内閣支持率について何かいえるだろうか。実は何もいえないのである。」という。標本が母集団から無作為抽出されていない場合、母集団の推測統計ができない。

言い換えると、標準得点は作成されているが、その標準得点からは母集団での位置は推定できない。具体的には、標準得点が 70 点であっても、5 名に 1 名程度のありふれた得点なのか、100 名に 1 名程度のまれな高得点なのか、解釈不能である。

いくつかの質問紙法を以下にまとめておこう。

- CF 式 (千葉大式) 性格検査の標準化は、高校生 (男性 167 名、女性 192 名)、大学生 (男性 178 名、女性 181 名)、社会人 (男性 146 名、女性 93 名) の計 957 名である¹⁴⁾。しかし、サンプリング手続きは不明である。したがって、機会サンプリングである。この方法では日本人全体を代表する標本を集めることは困難である。この性格検査は YG の改良版として企画されたが、現在、消滅状態である。
- TEG の第 2 版は、男性 3175 名 (平均年齢 36.1 ± 12.1 歳)、女性 2462 名 (平均年齢 32.9 ± 12.0 歳) の計 5637 名で標準化された。年齢範囲も広く、職業は、製造職、事務職、管理職、営業職、学生、主婦などである。調査地域も 19 都道府県に及んでいる (東京大学医学部心療内科 TEG 研究会, 2006)¹⁵⁾。ただし、サンプリング手続きや被検者の内訳は公表されていない。したがって、機会サンプリングである。標本は大規模であるが、大きな標本を取りさえすればよいのではない。この標本も日本人全体を代表するものではなく、標準得点は利用できるが、その意味は不明確

である。

- FFPQ は大学生（男性 239 名、女性 691 名）で標準化された。被検者は七つの大学で心理学関係の授業を受講していた大学生であった。サンプリング手続きは不明である。したがって、機会サンプリングである。辻斉は、「大学生の標本が、パーソナリティに関して、母集団（日本に住み日本語を母国語とする人のデータ）に比べて特に偏っているとは考えにくい」（辻,1998, P.127）¹⁶⁾ というが、大学生は入試という関門を通過した集団で、すべて心理学関係の受講者で、切断正規分布をしている可能性が高い。しかも、情緒的に安定していて、精神障害者は少ない。標本は日本人全体から無作為抽出されていないので、日本人全体を母集団とすることはできない。図1は、大学生標本に基づいた精神分裂病分類不能型（DSM-III-R による診断）の MMPI-1 プロファイルの例である（村上・村上, 1992）。同じデータを、住民票からの多段階無作為抽出法による標準化基準（村上・村上, 1994）に基づいて作成すると図2のプロファイルになる。図1では、90点以上が尺度8と4に二つもあるが、多段階無作為抽出後の基準を用いると、80点かそれ以下まで低下する。したがって、FFPQの標準化でも高得点や低得点が強調される結果となる。

健常者は日本人の代表ではない。

健常者を意図的に収集した質問紙では GHQ（精神健康調査票）がある。ここでは精神科医が現在症診察（一時間近くかかる構造化面接）を行って健常者を抽出した。その目的は、得点の標準化ではなく、神経症者との判別である。

日本人全体を母集団とする場合、無作為抽出された標本は、精神的に不安定な人々も含まれるはずである。意図的に健常者のみを抽出すると、その標本は母集団を代表しなくなるので、標準化に用いるべきではない。

MMPI 新日本版の標準化は、作成委員会が、全国を8地方に分けて、1990年の国勢調査の人口数に比例した人数を、標準化の協力者190名に配分

した。次に、協力者190名が、国勢調査の結果に近づくように、被検者を意図的に選択した。したがって、機会サンプリングである。そこには「心身とも健常な被検者を選定すること」、「被検者が回答中は、原則としてその場面にいること」、「MMPIの標準的な教示に加えて、『どちらともいえない』は9個以下とするように指示すること」、「回答欄が冊子内に組み込まれている冊子(タイプA 質問票)」を使ったことが記載されている。

標準化データを収集する場合、実施法は標準的な実施法とほぼ同一でなければならない。ところが、MMPI 新日本版の場合、冊子に回答欄が印刷されている物を使っていて、テスト実施者はそれを常に目視している。観察されている場合は異常性のある回答がしづらいので、建前的回答が増えてしまうだろう。このような回答様式は、知的レベルの落ちた被検者などに向いているだけである。

標準化資料の収集後、「各年齢段階、職業、教育歴への標本の配分比率が平成2年の国政調査における構成比となるべく近くなるように調整した」とあり、最終的に男性500名、女性522名となった¹⁷⁾。

この標本抽出法では、日本人全体を母集団と設定することは困難である。標本は、「心身ともに健常な被検者」に限定されているが、その選定基準は不明確で、全国の協力者に委ねられた。そのため、大学生標本と同様か、それ以上に健全な標本に偏った可能性がある。

収集された標本が日本人全体から無作為抽出に近い方法で抽出されている場合、年齢段階、職業、教育歴などは自動的に国政調査の構成比と一致する。たとえば、住民票からの多段階無作為抽出法で得た MMPI-1 の標本の場合は、教育歴や職業などの構成は、国勢調査の結果とほぼ一致していたので、多段階抽出法は無作為抽出に近い方法であったと確認できる。しかし、MMPI 新日本版の標本は、標準化資料の収集後に「各年齢段階、職業、教育歴への標本の配分比率が平成2年の国政調査における構成比となるべく近くなるように調整した」とある。つまり、日本人全体から無作為に抽出された標本でなかったため、それを人為的に調整したことが明らかである。

MMPI 新日本版の標本は健常者に偏っているため分散が小さく、Z得点を利用しているため、逆に

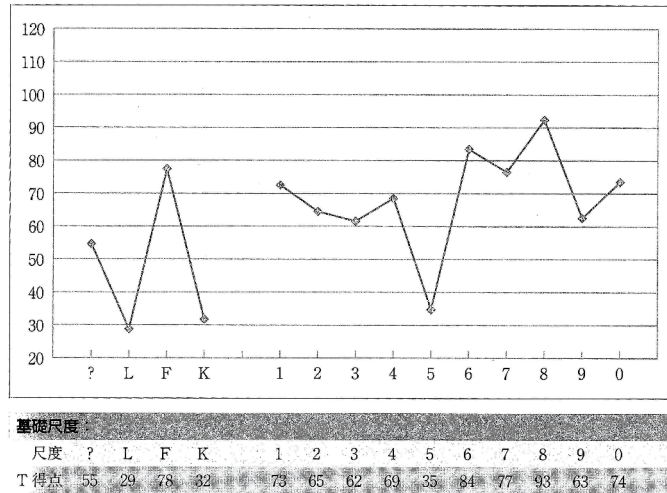


図 3: MMPI 新日本版のプロファイル例。野呂・他（2011, P.125 より引用）では不安障害等の診断名であるが、68 コードなので、統合失調症や妄想性障害ではないか。

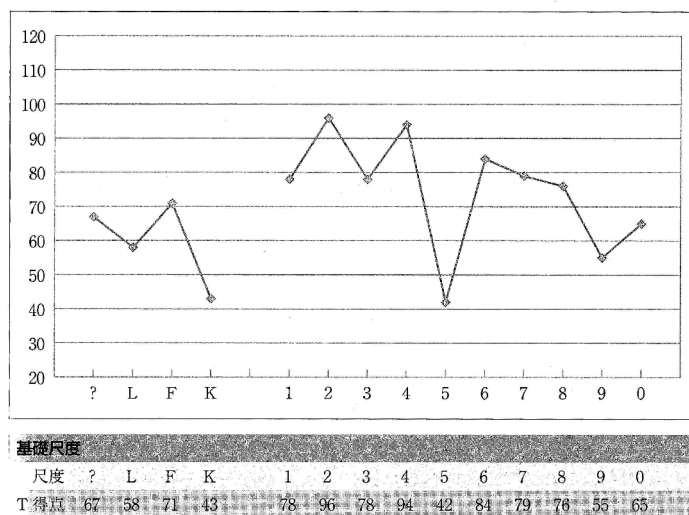


図 4: MMPI 新日本版のプロファイル例。野呂・他（2011, P.115 より引用）で外傷後ストレス障害とあるが、24 コードなので、パーソナリティ障害ではないか。

標準得点の分散は大きくなる。言い換えると、高得点と低得点が頻繁に得られるはずである。標準化データは一つしかないので、MMPI-1のように大学生標本と多段階無作為抽出法による標本の結果を比較できないが、野呂・他(2011)¹⁸⁾を見ると、高得点と低得点が頻繁に得られることが分かる。

図3は、不安障害、解離性障害、身体表現性障害という診断だが、MMPI プロファイルを見ると尺度8と6が極端な高得点である。86コードの場合は、統合失調症や妄想性障害の可能性が大きい。プロファイルで明らかのように、もし、正規分布をしているなら、80点と90点が同時に得られる理論確率はほぼゼロである。10万人に1人程度の希な事例であろうか。実は、不安障害などはありふれていて、10名に1名ほど見られても不思議ではない。逆に表現すると、MMPI 新日本版では高得点と低得点が大きく強調されているので、正常に近い人でも精神障害と判定してしまう恐れがある。

図4は、外傷後ストレス障害という診断だが、尺度2と4が極端な高得点である。24コードの場合は、パーソナリティ障害や物質関連障害の可能性が高い。事例の説明によると暴力行為は目立つが、嗜好性などは記述されていない。したがってパーソナリティ障害であろう。このプロファイルでも90点を超える尺度が二つもあり、理論確率は、ゼロの値である。したがって、プロファイルの尺度得点が正しいのではなく、標準化データが健常者に偏っているために、人為的に極端な高得点を得られたと思われる。

5～6章の13症例のMMPI 基本プロファイルから、最小値、もしくは、最大値を代表値として取り出した。すると、29、60、72、75、84、93、96、100、110、110、120、130、149となった。この内、1例は治療終結まぎわであった。50点から3 σ 以上離れている得点は、13例中9例にも及ぶ。このような超高得点は、理論確率がほぼゼロである。非常にまれな患者だけが集まったと解釈するのは困難で、標準化に使った標本が不適切と考えるべきである。

この本の大きな欠点は診断基準が不明で、診断名の妥当性が乏しいことである。事例の診断名と、プロファイル・タイプから推測される診断名が明らかに矛盾している。しかし、なぜ、そうなるのかという説明はない。第一筆者は金剛出版から書評を頼ま

れた関係で、「コードと矛盾する解釈は果たして正しいのだろうか。診断や治療に関する情報提供が不十分である」、「T 得点のレベルが異常に高く、90点から120点を超える事例が多いことにも注意すべきである」などと書いた。MMPI 新日本版では、極端な高得点を得られるため、健常者でも標準得点が上昇しがちであり、精神障害者と誤って診断する可能性が大きい。

正規化 T 得点による質問紙

主要5因子性格検査もMMPI-1と同様に、住民票からの多段階無作為抽出法を採用した。全国を北海道、東北、関東・信越、中部、関西、中国・四国、九州・沖縄の7ブロックに分け、23都道府県、46区市町村を抽出した。四国、沖縄は電話帳からの無作為抽出であった。4100名に郵送法で依頼し、有効回答数は1166名であった。MMPI-1、MINI、MINI-124に準拠し、青年期(12～22歳、男性123名、女性131名)、成人前期(23～39歳、男性130名、女性183名)、成人中期(40～59歳、男性159名、女性155名)、成人後期(60歳以上、男性163名、女性122名)に分割して標準化を行った¹⁹⁾。

標本抽出は、多段階無作為抽出法であるが、標準化では正規化 T 得点法が適用された。すでに説明したように、パーセンタイル順位を標準得点に変換するので、標準得点とパーセンタイル順位が一对一对応する。得点解釈に一意性があるので、妥当性も確保しやすい。正規化 T 得点法を用いて標準化を行ったのは、日本では主要5因子性格検査のみである。

アメリカの MMPI の標準化

アメリカのMMPIの標準化にも問題は多い。国柄のせいか、層化多段階抽出法に基づいて、郵送法で標本を求めるにしても、日本のように協力する人は少ない。郵便の信頼性も乏しい。

ミネソタ基準

ハサウェイとブリッグス²⁰⁾によるミネソタの基準集団から選抜された集団がもっとも広く用いられた。男性226名、女性315名で、平均年齢はそれぞれ33.1歳、33.9歳である。学歴がほぼ一定にそ

ろえられている。ただし、? (不応答)、L、F (頻度)、K の妥当性尺度には、任意の得点が割り当てられている。これ以外にも、ミネソタの高齢者基準集団やノースカロライナの基準集団がある。

MMPI は基準関連的尺度構成を行ったために、ターゲットとなる確定診断された基準群と対照群が必要であった。暫定的な質問紙を二つの群に実施し、回答差のある項目を集めて尺度を作成した。手間のかかる方法である。ハサウェイは MMPI の開発に 10～20 年を費やした。そのため、ハサウェイにはきちんと標本抽出をするエネルギーはなかったのだろう。

標準化のための標本はミネソタ大学病院を訪れた健常者が多く、アメリカ人全体という母集団を代表する標本ではない。そのために、極端な高得点も得られやすかった。被検者が思春期の場合は、得点を割り引いて解釈せざるをえなかった。言い換えると、MMPI の標準化はかなり不完全で、修正が必要な項目もあった。

正規化 T 得点による MMPI

1984 年、メイヨウ・クリニックのコリガンら²¹⁾は、電話帳を用いて人口規模別に世帯を無作為抽出し、手紙により協力を申し込んだ後、電話で家族構成や年齢、教育歴、心身の疾病、調査への同意の有無などを調査した。同意が得られた 1442 世帯 2658 名から、心身のハンディキャップのある人や辞退者などを除外した。そして、同意がえられた 1711 名中、MMPI を最後まで完成したのは 1408 名だった。次に、1980 年合衆国国勢調査の白人成人の人口構成の基準に当てはまるように選択し、最終的に標準化に使用したのは、計 640 名 (男性 335 名、女性 305 名) であった。

結局、最初に同意した被検者の 24.1% が標準化の基準となった。年齢ではほぼ国勢調査の基準を満たしたが、教育年齢は全体的に非黒人基準より高くなった。K 修正付きと K 修正なしの二つの表がある。標準化は基本尺度しかなされなかったが、ウィギンズ内容尺度などの標準化表は後に発表された。

コリガンらの正規化 T 得点は、母集団でのパーセンタイル順位と一対一に対応するので、解釈の妥当性は保ちやすい。しかし、問題は、ミネソタ基準が Z 得点による標準得点で、MMPI のプロファイ

ルタイプ研究との整合性がとれなくなった。それで、今までのプロファイルタイプの研究を利用するためには、プロファイルの変換表を作成せざるをえなくなった。

1982 年に、ミネソタ大学出版会はサイコロジカル・コーポレーションから MMPI のライセンスを取りもどした。ハサウェイが 1984 年に死去したため、このライセンスは著作権を含むものと推測される。ミネソタ大学出版会は、MMPI の改訂を目指していた。そのため、コリガンたちの正規化 T 得点による MMPI は広く使われるに至らなかった。

MMPI-2 の一様 T 得点

MMPI が病院臨床以外で使われるようになると、質問項目に含まれていた攻撃的・差別的表現が問題になった。そこで、MMPI-2 は 550 項目のうち 141 項目の攻撃的・差別的表現の手直しを行い、重複項目 16 と基本尺度の 13 項目とその他 77 項目を廃止し、タイプ A 行動や薬物使用等の内容尺度のため 107 項目を追加して全 567 項目となった。伝統的な 13 の基本尺度の他、三つの妥当性指標、15 の内容尺度、15 の追加尺度がある²²⁾。MMPI-2 の標準化は、カリフォルニア、ミネソタなど 7 州からアメリカの国勢調査に合うように 2600 名の標本を収集した。標準化は一様 T 得点というものを提案した。最初に Z 得点を算出し、基本尺度と K 修正された尺度、計 16 の分布ごとに Z 得点と対応するパーセンタイル順位を求め、それらを 16 の分布の間で平均した。一様 T 得点は分布の歪みを平均化した。従来プロファイル・タイプの研究を利用するためには、線型変換の T 得点を別に算出する手間が必要になった。

結局、MMPI-2 は攻撃的・差別的表現のある項目を手直したが、ハサウェイが目指した臨床的診断の精度を上げるための基準関連的尺度構成は行わず、因子分析等による内容尺度を増やしただけであった。

MMPI-2-RF/MMPI-3

MMPI-2 は少なくとも長すぎた。また、臨床尺度の妥当性も低下したようだ。ベン・ポラスと

テレゲンは、2008年、MMPI-2から338項目を抽出してMMPI-2-RFを構成した。MMPI-2-RFは、MMPI-2の9つの妥当性尺度、9つのRC(意気消沈)尺度、5つのPSY-5(パーソナリティ障害)尺度や特殊尺度などから構成されている。

これを手直しして、335項目にしたのがMMPI-3で、2020年に出版された²³⁾。MMPI-3はアメリカの人口構成に合わせて男性810名、女性810名の標本を集めた。妥当性尺度群11、上位尺度3、再構成尺度8、身体/認知・内面化尺度14、外面化・対人尺度12、性格異常尺度5である。プロファイル・タイプによる診断情報は一切利用できない。ハサウェイの臨床尺度はすべてなくなった。言い換えると、ハサウェイの古典的MMPIはこれで消滅した。

MMPIの凋落

MMPIの利用者はどうなったのか。GoogleによるTrend分析を図5に示す。グラフは2004年からの検索頻度を表す。現実の使用頻度は意味しないが、ある程度は反映されているだろう。古典的MMPIは流通していないが、検索頻度が比較的高い。これは教科書などに掲載されているからであろう。

1970～1980年代になり、古典的MMPIに対する批判は大きくなった。それには複数の要因がある。一つ目は古典的MMPIがミネソタ大学病院を訪れた健常者集団で標準化されていたため、標本の分散が小さく、そのために標準得点の分散が大きくなったと推測される。病院臨床以外の場で用いられると、その弊害が目立ってくる。得点解釈の妥当性も低下しただろう。思春期の被検者では極端な高得点が得

られやすかったので、割り引いて解釈する慣行も生まれた。

二つ目は質問項目に攻撃的・差別的表現があり、病院臨床以外で用いられると、大きな批判が上がったためである。そのため、古典的MMPIは裁判関係など、限定された領域以外での使用が難しくなった。これに対応してMMPIの項目の改良に取り組んだのがブッチャラのMMPI-2である。MMPI-2では攻撃的・差別的表現が書き換えられたが、基準関連の尺度構成を行ってMMPIの伝統的な臨床尺度の妥当性を高めたり、確認することは行われなかった。そのため、臨床尺度の妥当性は低下したと思われる。MMPI-2は一貫して関心を持たれなかった。

三つ目は、1980年にアメリカ精神医学界が「精神障害の診断のための手引き」(DSM-III)を正式に採用したことである。それまでは力動精神医学が主流で、精神科医は診断する基準がなかった。そこに、基準関連的に尺度構成した古典的MMPIの存在意義があった。ところが、DSM-IIIはクレペリン学派の記述主義を採用していた。患者の症状を手引きの症状リストと照合すると、ほとんどの場合、正確に診断できた。家庭医や非医療者もこの診断基準を利用するようになった。すなわち、患者に大規模な質問紙を実施して、それを診断に利用する必要がなくなった。この診断基準は1987年にDSM-III-R、1994年にDSM-IV、2013年にDSM-5と順次改訂され、現在に至っている。

MMPI-3はベン・ポラスとテレゲンらによる新規に作成された尺度がほとんどである。古典的MMPIの情報はすべて利用できない。GoogleTrendの分析によると、関心を持つ人は非常に少ない。

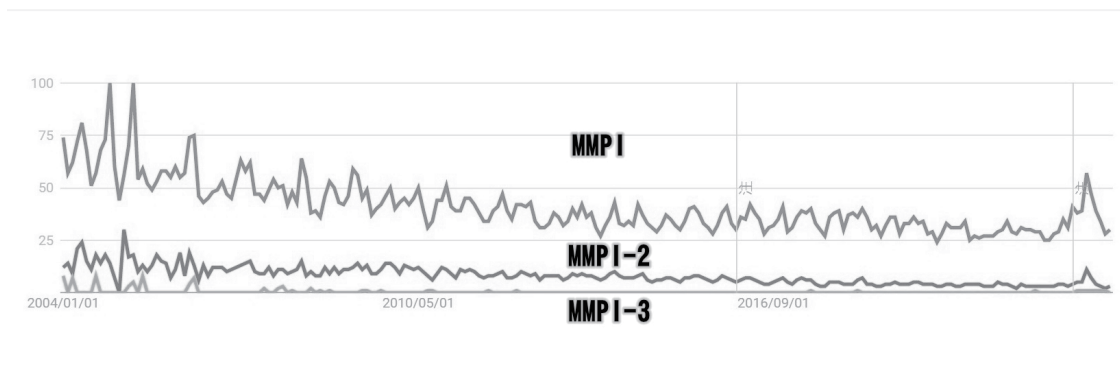


図5: GoogleによるTrend分析。

文献

- 1) 小嶋雅代・古川壽亮（訳著）2003 日本版 BDI-II について．日本版 BDI-II 手引き．日本文化科学社．
- 2) ゴールドバーグ, D.P. (著) 中川泰彬・大坊郁夫（訳著）1981 質問紙法による精神疾患の発見．－精神・神経症症状の診断法および評価法－．国立精神衛生研究所モノグラフ．(Goldberg,D. P. 1972 *The detection of psychiatric illness by questionnaire*. London: Oxford University Press) ．
- 3) ホーエル, P. G. (著) 浅井晃・村上正康（訳）1978 入門数理統計学．培風館．
- 4) ホーエル, P. G. (著) 浅井晃・村上正康（訳）1981 初等統計学．培風館．
- 5) 盛山和夫 2004 社会調査法入門．有斐閣．
- 6) 池田央 1994 現代テスト理論．池田書店．
- 7) Kline, P. 1986 A handbook of test construction. New York: Methuen.
- 8) 統計 Web <https://bellcurve.jp/statistics/blog/18075.html>
- 9) 川上憲人 2016 精神疾患の有病率等に関する大規模疫学調査研究：世界精神保健日本調査セカンド．総合研究報告書 <http://wmhj2.jp/WMHJ2-2016R.pdf>
- 10) 村上宣寛・村上千恵子 1992 コンピュータ心理診断法．－MINI, MMPI-1 自動診断システムへの招待－．学芸図書
- 11) 村上千恵子・村上宣寛 1996 MMPI-1, MINI, MINI-124 の世代別標準化研究, 性格心理学研究, 4, 23-37.
- 12) 標準正規分布の計算サイト <https://keisan.casio.jp/exec/system/1174204351>
- 13) サール, A. (著) 宮本聡介・渡邊真由美（訳）2005 心理学研究法入門．新曜社．
- 14) 柳井晴夫・柏木繁男・国生理枝子 1987 プロマックス回転法による新性格検査の作成について．I. 心理学研究, 58,158-165.
- 15) 東京大学医学部心療内科 TEG 研究会（編）2006 新版 TEG2 解説とエゴグラム・パターン．金子書房．
- 16) 辻平治郎（編）1998 5 因子性格検査の理論と実際：こころをはかる 5 つのものさし．北大路書房．
- 17) MMPI 新日本版研究会（編）1993 新日本版 MMPI マニュアル．三京房
- 18) 野呂浩史・荒川和歌子・井出正吾 2011 わかりやすい MMPI 活用ハンドブック－施行から臨床応用まで－ 金剛出版．
- 19) 村上宣寛・村上千恵子 1999 主要 5 因子性格検査の世代別標準化．性格心理学研究, 8, 32-42.
- 20) Hathaway, S.R., & Briggs, P.F. 1957 Some normative data on new MMPI scales. *Journal of Clinical Psychology*, 13, 364-368.
- 21) Colligan, R.C., Osborne, D., Swenson, W.M., & Offord, K.P. 1983 *The MMPI: A contemporary normative study*. New York: Praeger Publishers.
- 22) Butcher, J.N., Graham, J.R., Williams, C.L., & Ben-Porath, Y.S. 1990 *Development and use of the MMPI-2 content scales*. University of Minnesota Press: Minneapolis.
- 23) <https://www.pearsonassessments.com/store/usassessments/en/Store/Professional-Assessments/Personality-%26-Biopsychosocial/Minnesota-Multiphasic-Personality-Inventory-3/p/P100000004.html>

受付年月日（2022/ 9/15）

受理年月日（2022/12/20）