

# 性格検査への項目反応モデルの適用

教育情報科学研究室 井上 俊 哉

## An Application of Item Response Model to the Personality Tests

Shunya INOUE

It has been known that Item Response Models are very useful to be applied to practical tests for their features. And researchers have done lots of work about the application of these models in the fields of ability or aptitude tests. But so far, there are few studies on personality tests. This research tries to apply 2 parameter logistic model, one of the Item Response Models, to the scale of sociability and depressiveness. First, item characteristic parameters were estimated in each scale. And by using information function, accuracy of each scale is evaluated as the function of subject's score. Then, at the limited level of character, the items included in each scale are reduced without spoiling the accuracy of estimated scores.

### 目 次

- I 問題
- II 方法
  - (適用の対象)
  - (適用)
  - (質問紙)
  - (被験者)
- III 結果と考察
  - 1 向性の尺度
    - a 項目パラメタ値
    - b 情報関数
    - c 短縮版の作成と評価
    - d 再検査データの検討
  - 2 抑うつ性の尺度
    - a 項目パラメタ値
    - b 情報関数
    - c 短縮版の作成と評価
    - d 再検査データの検討
- IV 要約
- 注
- 文献
- 付記

### I 問 題

現在、性格検査（質問紙）への回答の処理は、多くは

古典テスト・モデルとよばれるモデルに従って行なわれている。このモデルによる結果の処理を①精度の評価、②得点化の2点からみると、「信頼性係数」による精度の評価、「正反応項目の単純加算」による得点化ということになる。例えば、矢田部ギルフォード性格検査（以下 YG と略す）では、あらかじめ集めた項目を、因子分析をもとに 13 の下位尺度（各 10 項目）に分けたうえで、各尺度ごとの信頼性係数がもとめられる。そして尺度の正反応数とその尺度の得点とされる（辻岡1957）。しかし以上のような精度の評価および得点化には、いくつかの欠点がある。まず、信頼性係数は、以下のような短所を持つ。①被験者集団に依存する。係数は被験者集団が異質であるほど高くなる。逆に、真の得点が等しい被験者ばかりならば、係数は0になってしまう、②用いられた検査に依存する。検査全体で定義され、項目を少しでも入れ換えれば、はじめから計算しなおさなければならない。③信頼性係数は被験者集団ごとに一つの値がもとまるだけである。実際の検査においては、満点に近い被験者では天井効果、および0点に近い被験者では床効果が現われうる。ところが、信頼性係数は精度の個人間差異を無視している。

つぎに、正反応項目の単純加算による得点は、尺度に含まれる項目数および項目の強度に依存する（ここで、項目の正反応率と関係する概念として項目強度という言葉を用いることにする。正反応率の高い項目を強度の小さい項目、正反応率の低い項目を強度の大きい項目とす

る。すなわち正反応されにくい項目ほど強度は大きい)。項目数が多くなれば得点は大きくなりがちである。また尺度中に項目強度の大きい項目が揃っていれば、得点は小さくなりがちであるし、逆に強度の小さい項目が多ければ得点は高くなる。これは、各項目の強度の情報が得点に反映されていないからである。従って、異なる項目から構成された尺度間の得点を比較することはできない。

以上の欠点のため、性格検査の処理に当って、当該の性格検査の精度を被験者ごとに評価するのが困難となるし、異なる被験者どうしを比較するには、すべての被験者に同じ項目を実施しなければならなくなる。しかし一方、すべての被験者に同じ項目を実施した場合、測定される特性に関して高い値をもつ被験者は強度の小さい項目には正反応する確率が非常に高く、それらの項目は検査の精度にほとんど貢献しない(逆に測定される特性の低い被験者にとっては強度の大きい項目が精度に寄与しない)という問題が生じる。例えば、MMPI (Minnesota multiphasic personality inventory) は、精神病患者群と正常者群とをよく弁別すると言われる。一方、YG は、日常的な性格検査として一般によく用いられている。どちらの検査にも抑うつ尺度が含まれ、両尺度には、よく似ている項目もいくつかある。両尺度が抑うつ性という共通の心理特性を測定しているとみなしうるとしても、両者の結果を直接に比較することはできない。まず項目数がまったく異なる。さらに、うつ状態の強い個人についての精度を重視する MMPI の方が YG よりも強度の大きい項目が多いと考えられる。もしそうならば、仮に項目数を揃えても、抑うつ性の低い被験者が受験した場合を考えるならば、YG の抑うつ性得点が高くなることが予想される。また、抑うつ性の高い被験者が YG の抑うつ性尺度項目を受験したならば、天井効果によって精度が落ちることも予想される。

最近になって、この古典テスト・モデルにかわる項目反応モデル(芝 1978, Hambleton & Swaminathan 1985ほか)が注目を集めている。項目反応モデルは、被験者の項目への正反応の確率を、潜在的な心理特性の関数であらわすモデルである。このモデルのもちいる関数は、項目特性関数とよばれ、被験者特性値のパラメータと項目強度のパラメータを含む。両パラメータの間には、被験者の特性値が同じならば強度の小さい項目ほど正反応される確率が高く、項目の強度が同じならば特性値の大きい被験者ほど正反応する確率が高いという関係がある。そして、多数の項目への大勢の被験者のデータがあれば、統計的推定によって両方のパラメータをもとめることができる。また各項目のパラメータが明らかなき、これをも

とに被験者の特性値をもとめることができる。項目反応モデルには、いくつもの種類があるが、そのうち2パラメータ・ロジスティック・モデルの項目特性関数は

$$P_j(\theta) = \frac{1}{1 + \exp[-1.7a_j(\theta - b_j)]} \quad (1)$$

である。ここで $\theta$ は被験者の特性値を、 $a_j$ と $b_j$ とは項目 $j$ のパラメータ値をあらわす。 $b$ は能力検査では通常、項目困難度のパラメータとよばれるが、ここでは、より一般的に、項目強度のパラメータとよぶことにする。 $a$ は項目識別力とよばれる。 $P_j(\theta)$ は、項目 $j$ への正反応率である。

従来の古典テスト・モデルにかわり、項目反応モデルを性格検査に適用することができれば、モデルの特徴を生かして、様々な利点が得られるであろう。①特性値による表現は、各項目と心理特性との関数さえ特定されているならば、どの項目が用いられるかに依存しない。そこで、相互に比較したい被験者に同じ項目を実施しなくても、両者の結果を比較できる。具体的に言うと、例えば、MMPIとYGの抑うつ尺度に含まれる項目のパラメータ値を、共通の抑うつ性次元上に求めることができるならば、両尺度の受験者の結果を直接比較できることになる。②特性値の尺度は、多数の項目に対する多数の被験者のデータから、モデルにもとづいて構成されるので、何ら母集団を想定することなく、多くの被験者からの反応をデータとして蓄積してゆけばよい。これにより、母集団の定義、標本抽出の困難を回避できる。③さらに、項目反応モデルでは、検査の精度が情報関数とよばれる特性値の関数で与えられるという特徴をもつ。情報関数は、1項目ごとに

$$I(\theta) = 1.7^2 a_j^2 P_j(\theta) [1 - P_j(\theta)] \quad (2)$$

によって与えられる。式からわかるように、項目の情報関数は項目パラメータが既知であればもとめることができる。検査全体の情報関数は、項目の情報関数の単純和として計算することができる。すなわち、 $n$ 項目の検査では

$$I(\theta) = 1.7^2 \sum_{j=1}^n a_j^2 P_j(\theta) [1 - P_j(\theta)] \quad (3)$$

この情報関数により、被験者の特性値に応じて、その精度を評価できる。つまり、検査が個々の被験者の特性値をどのくらい正確に測定しているかを判断することができる。④また、特性値が被験者集団に依存しないのと同様に情報量も被験者集団に依存しない。このような情報関数の特徴から、項目パラメータが既知の項目が多数あれば、これをもとに、望みどおりの精度をもつ検査を作ることが可能になる(Lord 1977)。例えば、うつ状態の重

い被験者にのみ関心をもって検査を構成するならば、特性値（抑うつ性）の大きい範囲での情報関数だけを考慮して項目を集めれば、必要最小限の項目で精度を落とさずに特性を測定できる。

このようなすぐれた特徴のため、古典モデルに比してのモデルの複雑さ、当てはまりのための条件の厳しさにもかかわらず、さまざまな実用の試みがなされてきている（例えば、芝 1978をはじめとする一連の語彙理解尺度への適用研究）が、多くは能力検査や学力検査への適用であり、それらの分野での研究の進展に比べて、性格特性の分野への適用はかならずしも多くはない。そこで、本研究では、性格検査にこの項目反応モデルを適用することを考える。なお、いくつもある項目反応モデルのうちで、2パラメタ・ロジスティック・モデルを用いる。

## II 方 法

### A 適用の対象

項目反応モデルでは、潜在特性の一次元性を仮定している。そこで、適用を試みる特性を選ぶにあたっては、既存の代表的な質問紙項目に因子分析を行なった研究を参考にして、それらの研究で潜在特性の一次元性が認められていることが前提になった。条件を満たす特性のうちから、代表的な性格特性として向性が選ばれ、さらに、いわゆる性格とはいくらか異なる次元と考えられる抑うつ性にも適用が試みられることになった。

（適用）

まず、最尤推定法によるコンピュータ・プログラム LOGIST (Wood, Wingersky & Lord 1976) を使って各項目のパラメタ値、すなわち識別力と強度が推定された。識別力が大きい項目を吟味することによって、尺度により測定される特性の内容の輪郭が得られると考えられる。また、識別力が小さすぎる項目は後の分析では使えないであろう。強度に関しては、強度の大きい項目は、特性値の高い被験者でないと正反応されにくい項目、逆に強度が小さい項目は、特性値の低い被験者でも正反応しやすい項目であると言える。ここで得られた項目パラメタ値は、以下の分析に利用される。

つぎに、向性、抑うつ性の両尺度について、その情報関数を求める。これにより、両尺度がそれぞれの特性のどのレベルで精度の高い尺度であるか、特性の高低によって、精度がどのように変化するかが分かる。

さらに、関心に応じて、被験者特性値の範囲を限定したうえで、情報関数と項目特性値を用いて、定められた範囲内で精度を落とさずに、両尺度の短縮判を作成する

ことを試みる。

また、推定された項目パラメタ値を利用して、再検査データが得られた被験者について、1回目、および2回目の特性値を、別々に推定して<sup>1)</sup>、両者を2次元座標上にプロットする。データがモデルに当てはまっていれば、プロットは、傾き1の直線上にならぶはずである。

以上により、性格検査への項目反応モデルの適用の可能性と利点を吟味する。

（質問紙） 向性測定のための質問項目には、日本版 MBTI (加茂ほか 1965ほか) から31項目が選ばれた。それらの項目を YG の各下位尺度の項目と照合したところ、選ばれた31項目は、YG の3つの下位尺度(「社会的向性」、「支配性」、「のんきさ」)にまたがる内容領域をもつと考えられた。また、ここで選んだ31項目のうち、22項目に因子分析を行なった研究(加茂ほか 1965)によると、第一固有値(8.92)と第二固有値(1.24)以下の差がきわめて大きい。このことから、向性尺度の一次元性は、ほぼ問題ないと考えられる。

抑うつ性測定のための質問項目については、MMPI の抑うつ尺度の項目から集めることも考えられたが、Comrey (1957) の因子分析研究の結果によると、この尺度は、雑多な特性に関する項目を含むと判断された。そこで、先行研究<sup>2)</sup>によって一次元性が認められた TPI (Today Personality Inventory) からの34項目と YG の抑うつ尺度の10項目が、まず選ばれた。これらの項目は、精神病患者群と非患者群とをよく弁別する項目でもある。さらに項目の内容を吟味して、これをうつ病の症状と認められているもの、例えば、DSM-III(高橋ほか訳 1982) のうつ病の診断基準と対比すると、うつ病の症状の全域を含んでいない。そこで、うつ病の症状全般を含むことを目標に作成されたツングの質問紙の20項目(Zung 1965)もくわえられた。<sup>3)</sup> こうして集められた64項目のうちで、表面的にほとんど同じ項目は TPI に揃えて一つにまとめられ、最終的には58項目が選ばれた。

（被験者） 項目パラメタ値の推定、および他の分析に用いられた被験者は以下の通りである。

北海道立K高等学校全校の221名	
国立N大学での教育心理学受講生	
向性について	142名
抑うつ性について	141名
私立T女子大学での教育心理学受講生	
	107名
その他(高校生、主婦など)	
計 向性について	576名
抑うつ性について	575名

表 3.1.1 YG の3つの下位尺度に対応させた項目識別力の分布 (向性)

	識別力→									
	0	0.5		1.0		1.5		2.0		
S (社会的向性)		12	21	13	11	17	29	18	6	2
		23				28				
A (支配性—服従性)		20		16	8	5	15	1		14
				19			25			
				22			30			
R (のんきさ)		9	24	7	10	31	4		3	
		26		27						

注：枠内の数字は項目の番号

いずれも、1986年10月に調査が実施された。  
再検査の回答が得られた被験者は以下の通りである。

国立N大学での教育心理学受講生

向性について 159名

抑うつ性について 160名

1986年10月下旬と11月下旬

(いずれも間隔は4週間に調査が実施された

### III 結果と考察

#### A 向性の尺度

a 項目パラメタ値 最初に、コンピュータ・プログラム LOGIST を使って、被験者特性値 (以後  $\theta$ ) と項目パラメタ値とを同時に推定した。ここでは、項目パラメタ値について検討する。

まず、項目識別力について検討しよう。項目識別力は、各項目に対する反応と、この検査で測定される潜在特性との関連の大きさをあらわす。項目が選ばれたのちに、各項目が対応させられた YG の3つの下位尺度ごとに項目識別力の分布を示したのが表 3.1.1 である。これをみると、識別力の値が1より大きい6項目のうち4項目が「社会的向性」に分類されている。6項目からキーワードをさがすと、「社交」(項目2, 14), 「ひかえ目」(項目6, 14), 「静か」(項目3, 14) などである。対人的な、あるいは社交場面での積極性—消極性に関連のある項目と考えられる。一方、これらの項目を別にすれば、YG の下位尺度ごとに識別力の偏りはみられない。以上のことから、この向性尺度は、対人的な積極性—消極性をもっとも正確に反映しつつ、YG の各下位尺度が測定するものよりは一般的な向性を測定していると思われる。また、識別力は、コンピュータ・プログラムの設定する2.0を示した項目14を除くと、0.31 (項目12) から1.58 (項目

21) の範囲で分布している。これらの値を、識別力と因子負荷量  $r$  の関係をあらわす式

$$r = \frac{a}{\sqrt{1+a^2}} \quad (4)$$

を用いて換算すると、0.30 から 0.84 に相当する。YG の下位尺度ごとに因子分析を行なった場合、各尺度に用いられている項目からもっとも低い負荷量を拾うと、「社会的向性」で0.52, 「思考的向性」で0.34, 「抑うつ性」で0.54, 「のんきさ」で0.37 などである (辻岡 1957)。このことから、本尺度に含まれる各項目の識別力は、尺度を構成するのに妥当な大きさであると考えられる。

つぎに、項目強度について検討しよう。いくつかの項目について、表 3.1.2 に項目強度の推定値と正反応率 (内向方向への反応を正反応とする) を、図 3.1.1 に示した。項目31を除くと、正反応率が30%—40%であり、内向あるいは外向のいずれかの方向に偏った反応率を示す項目はない。向性  $\theta$  上で正反応率が50%になる位置をあらわす項目強度の値でみると、内向性の方向では項目31がとびぬけて大きく(1.41), 外向性の方向では、項目21以下、項目22, 8, 9 (-1.02) の順に外向性が強くなっている。外向性の高い被験者が項目31に内向方向の反応をする(「一日中机に向かっていなければならないこと」よりも、「はじめての人に会わなければならないこと」のほうが苦痛である) 確率は、かなり低い。逆に、内向性の高い被験者が項目9で外向方向の反応をする(「行動するまえにまず考える」のではなく「考えるよりまず行動する」) 確率は低いといえる。その他の項目は、 $-0.49 < b < 0.79$  の範囲におさまっている。

b 情報関数 得られた項目パラメタ値をもとに情報関数を求め、図 3.1.2 に示す。比較のために、抑うつ性、語彙理解尺度の両尺度についても同時に示した<sup>4)</sup>。横軸

表 3.1.2 向性尺度のいくつかの項目について項目強度のパラメータ値と正反応率

		項目強度 のパラメ ータ値	正反応率
31	あなたにとってより苦痛になるのは A 一日中机に向かっていなければならないこと B はじめての人に会わなければならないこと	1.41	0.19
12	休みのときには A ひとりで好きなことをしたいと思う B 友だちと一緒にすごしたいと思う	0.79	0.40
28	あなたは A 必要とあらば誰とでも楽に話せる B 特定の人と話す時や特別の場合だけ話題がみつかる	0.63	0.31
1	あなたにびったり感じられる言葉は A 親切 B もの静か	0.60	0.33
19	初対面の人を紹介されたときあなたは A 自分から話しかけるほうだ B 相手が話しかけてきてからこたえるほうだ	0.59	0.37
5	あなたにびったり感じられる言葉は A パーティ B 劇場	0.34	0.42
6	あなたにびったり感じられる言葉は A ひかえ目 B 話し好き	0.23	0.46
14	あなたは A ちょっとした社交家である B 静かでひかえ目なほうである	0.12	0.46
29	あなたは異性と A すぐ友だちになれるほうだ B なかなか友だちになれないほうだ	-0.07	0.52
24	話しがもつれたとき普通あなたは A 冗談を言ってふんいきをかえる B 数日たってもああ言えばよかったなどと考える。	-0.27	0.55
21	あなたのまわりのほとんどの人は A あなたの気持ちをよく理解している B あなたがうちあげたことのほかは知らない	-0.71	0.61
22	大勢の人の前で話すことは A 苦にならないほうだ B 苦手なほうだ	-0.84	0.68
9	あなたは A 考えるよりまず行動するほうだ B 行動するまえにまず考えるほうだ	-1.02	0.63

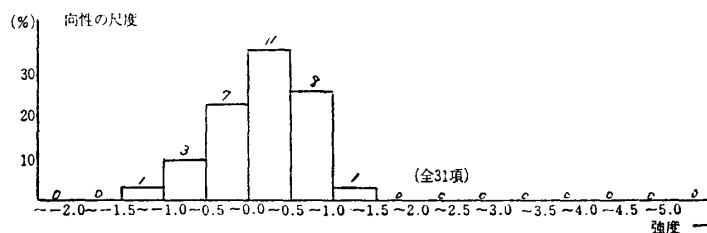
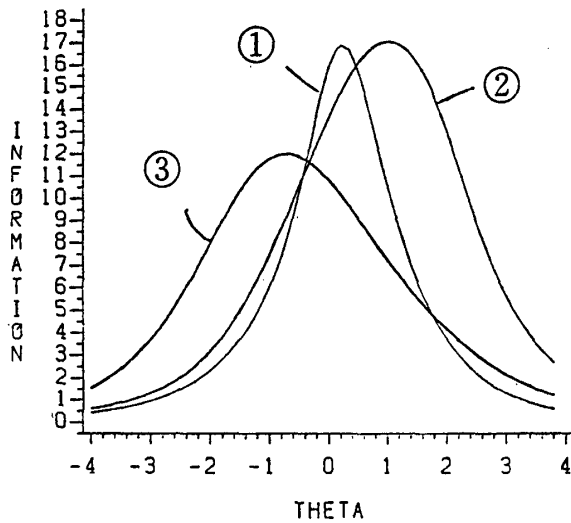


図 3.1.1 項目強度の分布 (向性)



①外向性尺度 ②抑うつ性尺度 ③語彙理解尺度  
 図 3.1.2 外向性尺度、抑うつ尺度、語彙理解尺度の情報関数

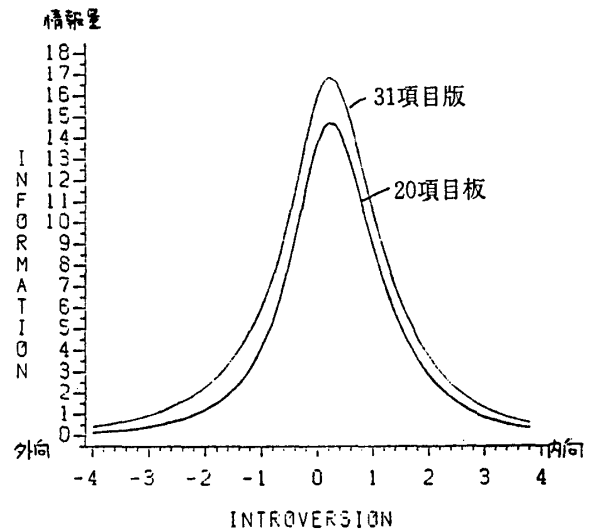


図 3.1.3 短縮版外向性尺度(20項目)の情報関数

が特性値、縦軸が情報関数の値(以下、情報量とよぶ)を表わす。右へゆくほど、内向性大、抑うつ性大、語彙理解力大である。外向性尺度では、特性値がほぼ0、あるいは、やや内向性よりで、ひととき情報量が大きいピークがあり、両端にゆくにしたがって情報量は急激に低下している。LOGISTでは、被験者集団の潜在特性値が平均0、分散1になるように定められているので、平均的な特性値をもつ被験者で精度が高いことになる。尺度の意味は特性値のデータを蓄積しながら、他のさまざまなデータと照合して明らかにしてゆかなければならないが、この情報関数の形は、特性値がピークの左側であるか右側であるかによって外向型の人(左側)と内向型の人(右側)とを分けている可能性もある。そうだとすると、この尺度は、内向型と外向型の人を弁別するのに高い精度をもつことになるが、確認は、今後の課題として残される。

c 短縮判の作成と評価 Lord (1977) が薦める項目情報関数を利用したテスト作成法は、以下のとおりである。

- ① 目標となるテスト情報関数の形を決める。
- ② 目標情報関数のもっとも大きいところで大きな情報量をもつ項目から選んでゆく。
- ③ 各項目をテストにくわえてから、その項目群によるテスト情報関数を計算する。
- ④ テスト情報関数が満足ゆく程度まで目標情報関数に近似したら、テストのための項目選定をやめる。

この方法は、内容の吟味が終わった。十分な数の項目ライブラリがある場合に適していると思われる。ここでは、項目数が少ないので、内容を主観的に考慮する余地

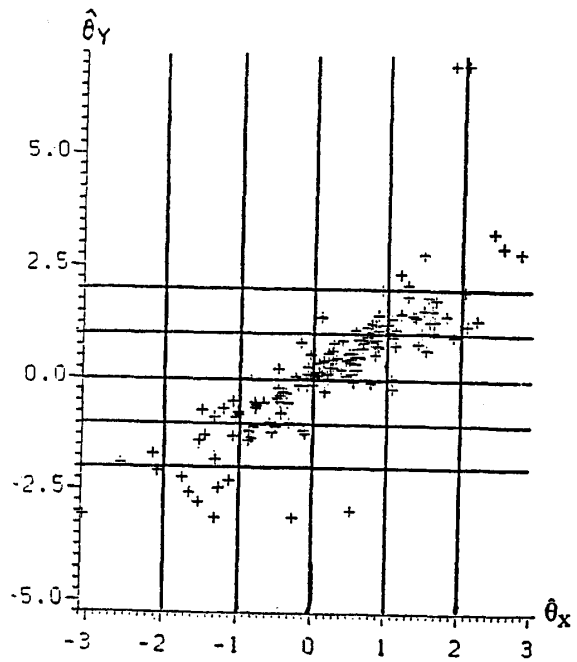


図 3.1.4 再検査データから推定された特性値( $\theta$ )の分布

- 注1 X軸が1回目のデータからの推定値  
 Y軸が2回目のデータからの推定値
- 注2 159名 相関係数  $r=0.79$   
 但し、1番外側のデータを取り除くと151名で  $r=0.86$

を残した、よりゆるやかな方法を用いる。その手順は、  
 ① できるだけ情報関数のピークの位置と高さを維持することを目標にする。これは、もとの情報関数の形から、外向性尺度の特徴は尺度上の平均値の近くで被験者を2群に弁別するのに高い精度を保つことであると判断されるからである。短縮版の項目数は20項目とする。  
 ② 所定の特性値における検査の情報関数とその値にお

ける項目情報関数の和であるという性質に注目して、特性値が0から1の近辺の各項目の情報量を調べる。

③上記の範囲で明らかに情報量の小さい6項目を除き、残る25項目から、内容の重複をできるだけ避けて、20項目を選ぶ。

④以上の手続きで選んだ20項目による情報関数が図3.1.3である。20項目版による情報関数は、そのピーク的位置、高さともに、ほぼ維持されていることがわかる。

d 再検査データの検討 再検査の間隔は4週間であり、対応するデータは等しい真の特性値( $\theta$ )からのデータと考えられる。モデルでは、所与の $\theta$ の推定値の標本分布が平均 $\theta$ 、分散 $1/I(\theta)$ の正規分布に従う(芝1978)ので、対応するデータを縦軸、横軸にとったプロットは、ほぼ傾き1の直線上にのり、特性上の情報量の小さいところで、ばらつきが大きくなるはずである。

図3.1.4は、プロットがよく一直線上にならんでいることを示している。ばらつきも中央付近で小さく、両端にゆくにしがって急激に大きくなってゆく。このことはデータがモデルに当てはまっていることの有力な証拠になると思われる。

### B 抑うつ性の尺度

a 項目パラメタ値 うつ病の症状の全域から項目を集めるために参考にしたDSM-IIIは、うつ病の症状を8つの範疇に分けている。そこで、用いた58項目をこの8範疇と「その他」の9つに分類して、各範疇に項目を対応させたい。項目識別力の分布を示した(表3.2.1)。表をみると、(1)の「食欲減退または体重減少」で低く、(3)の「精神運動性焦燥または抑制」、(4)の「日常活動への興味または喜びの喪失」で幾分高い。しかし、(3)と(4)は分類された項目も多く、識別力が高くなるのは、

ある意味で当然のことである。そこで、識別力が1以上の7項目を、識別力の大きい順に表3.2.2に挙げた。先の分類でいうと5つの範疇にわたっているが、あらためて項目をみると、脅威、不幸への予感、深い絶望といったものが感じられる。これらの項目から、ここで測定される抑うつ性の輪郭がとらえられるだろう。

次に、項目強度の値を検討しよう。表3.2.3に、いくつかの項目の強度と正反応率(抑うつ性の高い方向を正とする)を、図3.2.1に項目強度の分布を示した。項目強度 $b$ の原点は、特性上の任意の点にとることができるが、前にも触れたとおり、LOGISTでは、被験者集団の潜在特性値が平均0、分散1の正規分布になるように定められている。正の値の $b$ が多かったのは、ここでの被験者集団にとって相対的に抑うつ性の大きい項目が多かったことを意味する。この被験者集団にあって平均的な抑うつ性をもつ被験者が50%以上の確率で正反応する項目( $b < 0$ の項目)は、58項目中12項目にすぎない。正反応率でみても、20%以下の項目が16項目もある。また、 $b$ の値のちらばり方は、向性と比べて大きい。識別力が極端に小さくて当てはまりが疑われる数項目を除くと、項目強度は、 $-2.80 < b < 4.73$ の間となる。

項目パラメタの推定値をみて、特徴的なのは、質問項目選定の際に用いた質問紙ごとの違いである。まずツングの質問紙からの項目は、強度のもっとも大きい項目ともっとも小さい項目を含み、他は特性上にまんべんなく分布する。抑うつ性のさまざまな程度に対応するとも言えるが、識別力を見ると、極端に低い項目もあり、総じてあまり高くない。TPIからの項目の強度は、 $0.09 < b < 2.68$ の範囲で、比較的抑うつ性の強い項目群である。一方、識別力は、項目5の0.35を除けば、0.48から1.40の間で、これを因子負荷量に対応させれば、0.43から

表 3.2.1 DSM-IIIの8範疇を基本とする9範疇に対応させた項目識別力の分布(抑うつ性)

	0	0.5	1.0	1.5	2.0
1) 食欲減退または 体重減少	①②	20	③		
2) 不眠または 睡眠過多		④			
3) 精神運動性焦燥 または抑制		⑤ ⑥	50 ④ 27 9	54	13 18
4) 日常活動への 興味または喜びの喪失	⑦	⑧ ③ ⑨	26 57 6	⑩ ⑪	23
5) 気力の減退 易疲労性		35	⑫	14	11
6) 無価値感、 自責感、罪責感		⑬	51 34 21 55	46	
7) 思考力、注意集中力 の減退、決断困難	⑭	5	15 22 2	47	58
8) 死についてくり 返し考えること		⑮			39
9) その他	⑯ ⑰		43	38	
			53		

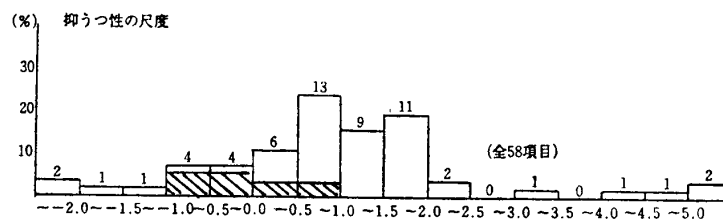
注1: 枠内の数字は項目の番号  
注2: ○で囲んだ項目はツング質問紙からの項目  
但し、3、19、25は他の質問紙と重複する項目

表 3.2.2 識別力が高い ( $a > 1$ ) 7項目 (順に)

18	ほとんど毎日 私をおびやかすようなことが 何か起こる
13	いつでも 何か不安の種になる
39	私は よく 死にたいと思う
23	いつまでも忘れられないような 強い失望をしがちである
38	ときどき 頭をしめつけられているような 感じがする
58	たびたび 物思いに沈むことがある
47	いやなことが 頭にこびりついて 離れないでいる

表 3.2.3 抑うつ性尺度のいくつかの項目の強度と正反応率

	強度のパラメタ値	正反応率
便秘で悩んでいる	4.73	0.11
夜眠れないので悩んでいる	3.32	0.08
ちかごろずっと体が弱っているような気がする	2.06	0.18
気が狂いはしないかと心配だ	1.85	0.10
落ち着かなくてじっとしてられない	1.71	0.29
私はよく死にたいと思う	1.68	0.11
私はかわいそうな人間だ	1.50	0.18
スランプにおちいて長い間何も手につかぬことがある	1.33	0.24
自分はごく限られた人としかが合わない	1.12	0.31
今の自分は本当の自分ではないように思う	1.00	0.28
物事を気楽にやれる	0.81	0.36
ふつうの人よりとくに神経質だ	0.69	0.40
へなへなと気がくじけてしまうことがある	0.30	0.42
私はよく憂うつになる	0.09	0.48
たびたび元気がなくなる	-0.40	0.61
ぼんやり考えこむせがある	-0.63	0.64
ときどき自分をつまらぬ人間だと思うことがある	-0.72	0.67
皆に役立つ人間と思われている	-1.75	0.73



グラフの上の数字は項目数。

図 3.2.1 項目強度の分布 (抑うつ性)

注: 斜線部は, YG からの項目

但し,  $b = 0.09, 0.66, 0.68$  の 3項目は TPI と重複

0.81になり, かなり高い識別力をもつと言える。TPI からの項目群は, ある程度以上抑うつ性の高い被験者に対して精度の高い項目と言えるだろう。YG からの項目群は, 識別力が比較的高く, TPI と同様に, ここで測定される一次元特性と相関の高い項目群である。TPI と異なるのは, その項目強度の分布で,  $-0.73 < b < 0.68$  の範囲にある。こちらの項目群は, 抑うつ性の小さい被験者で精度の高い項目と言えよう。

b 情報関数 図 3.2.2 の a が抑うつ性尺度全項目の情報関数である。特性値の平均は 0 であるから, 本研究の被験者集団にあっては抑うつ性が平均よりも高い場合に精度が高いことがわかる。特性値の大きさもつ意味がまだ明らかではないが, これは, 58項目のうちの34項目が正常者群とうつ病患者群をよく判別する TPI からの項目であったことを考えると, 特性上の精度の高い位置が両群の境界点に近いのではないかと予想される。



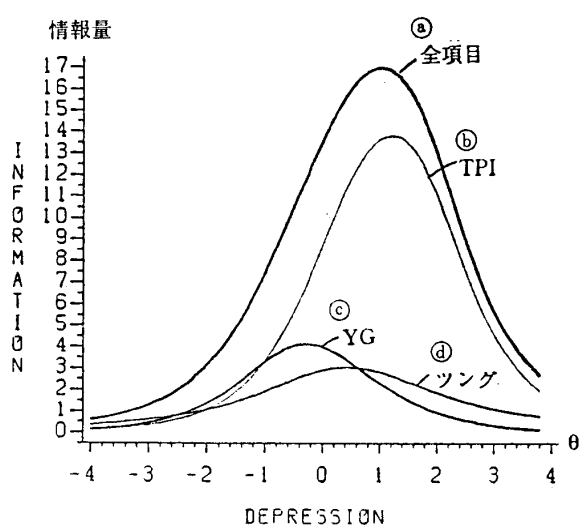


図 3.2.2 抑うつ尺度の情報関数

そこで、さらに、TPI、YG、ツングのそれぞれの質問紙ごとに情報関数をもとめて図3.2.2にb, c, dとして示した。この尺度において、YGからの項目群は抑うつ性次元の低い範囲で、TPIからの項目群はより高い範囲で精度が高いことが確認された。ツングからの項目群は両者の中間付近でいちばん情報量が大きい、総じてあまり高くないことがわかる。

c 短縮判の作成と評価 まず、うつ病の臨床経験の豊富な精神科医に、軽度のうつ病患者あるいはうつ病の疑われる人を想定したうえで、質問項目に回答してもらった。回答を他の被験者と同時に LOGIST で計算し

たところ、 $\theta=3.31$  という高い値が得られた。そこで、うつ病患者と非患者との抑うつ次元上の境界を、 $2 < \theta < 4$  の範囲に設定した。この範囲内で情報量を落とさないことを目標にして、項目数20への削減を試みた。手続きは向性的場合とほぼ同じで、①目標の範囲内で、各項目の情報関数を求め、これを項目削減の客観的基準とする。②選ぶ項目の内容にできるだけ幅をもたせる。そこでDSM-IIIのうつ病の診断基準を参考にした9つの範疇から、分類された項目数にできるだけ比例するように項目を選び、20項目を決めた(表3.2.4)。③この20項目による情報関数が図3.2.3である。抑うつ性の小さいところでは情報関数は激減しているが、目標として設定した範囲では、情報関数の減りかたが少ないことがわかる。

d 再検査データの検討 向性の節でも触れた通り、データがモデルに当てはまっていれば、プロットは、ほぼ傾き1の直線上にのり、抑うつ尺度上の情報関数の小さいところでは、ばらつきが大きくなるはずである。図3.2.4からプロットがほぼ直線上にならんでいることが示される。これを図3.2.3の情報関数と見比べると、情報量は $\theta$ が1からやや正方向よりで最大になっているがプロットのばらつきも、同じ付近で小さく、よく一直線にのっている。また負の方向の端ではプロットの散らばりが大きく、モデルからの予想の基本的な点は満たされていると言えるだろう。

表 3.2.4  $2 < \theta < 4$  の範囲の情報量を優先して選んだ20項目

1	私は いろいろな遊びやレクリエーションを 楽しんでやる。
4	夜眠れないので 悩んでいる。
6	私は かわいそうな人間だ。
7	私の頭は どうかしている。
9	不幸なことが 起こりはしないかと ひどく悩む。
11	何をやっても うまくできない。
18	ほとんど毎日 私をおびやかすようなことが 何か起こる。
26	私は ひとかどの人間になろうという希望を 失いそうになる。
34	私は すまないことばかり してきた人間だ。
35	ちかごろ ずっと体が弱っているような気がする。
37	自分のさびしい気持ちを 人に言えず 困っている。
38	ときどき 頭をしめつけられているような 感じがする。
39	私は よく 死にたいと思う。
43	いつも のどや胸が つかえているような気がする。
45	金や仕事のことで くよくよする。
48	食欲は 普通だ。
49	私の生き方は まちがっていた。
50	スランプにおちいって 長い間 何も手につかないことがある。
54	気が狂いはしないかと 心配だ。
57	今の自分は 本当の自分ではないように思う。

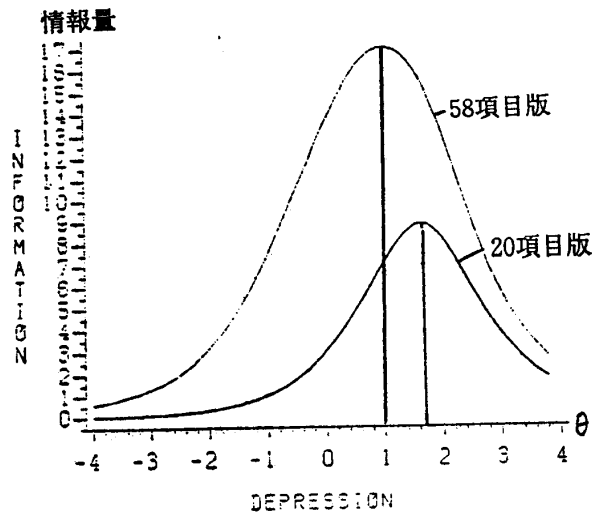


図 3.2.3 短縮版抑うつ性尺度 (20項目) の情報関数

#### IV 要 約

性格検査の結果の処理に、従来行なわれてきた古典テスト・モデルにかわって項目反応モデルを適用することを試みた。その結果、向性と抑うつ性の2つの特性に関して、モデルの適用によって、古典テスト・モデルでは

得ることのできなかった以下の利点を得られることが明らかになった。①項目パラメタ値が推定された。この値を利用することによって、実施する項目に依存しない被験者特性値をもとめることができる。②情報関数を利用することにより、検査の精度を個人ごとに評価できる。③重視する特性値の範囲での精度を落とさずに項目を減らすことができる。

一方、得られた結果の妥当性の吟味、異なる項目を受験した被験者の結果の等化の問題、モデルへのデータの当てはまりの細かい検討など、多くのことが課題として残された。

注1) ここで被験者特性値の推定に用いたコンピュータ・プログラムは、LOGIST ではない。項目パラメタの値を LOGIST で推定された値に固定して、 $\theta$  だけを2分法によって推定するものである。2分法のプログラムを使って最初の被験者全員の  $\theta$  を推定したところ、その推定値は LOGIST によるものとほとんど差がなかった。両プログラムによる推定値をプロットしたものが付図である。

注2) この研究は公刊されていない。放送教育開発センターの大塚雄作助教授と東京大学の鎌原雅彦助手の御好意でお借りすることができました。TPI の項目は必ずしも抑うつ尺度の項目ではない。この研究では、まず、TPI の500項目のうち正常者群とうつ病患者群を、よく判別する項目が判別分析によって選ばれた。つぎに、それらの項目にYGの項目をくわえた因子分析が行なわれ (52項

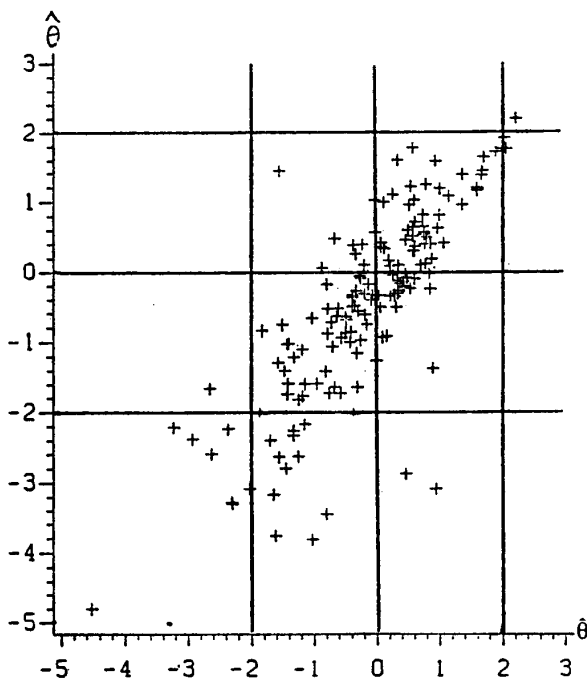
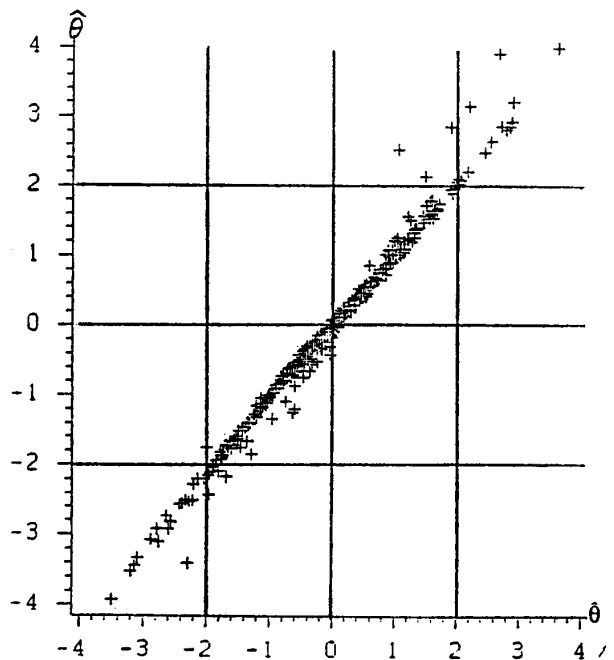


図 3.2.4 再検査データから推定された特性値 ( $\theta$ ) の分布

- 注1: X軸が1回目のデータからの推定値  
Y軸が2回目のデータからの推定値
- 注2: 160名 相関係数  $r=0.78$   
但し、1番外側のデータを取り除くと156名で  $r=0.82$



- 付図 2つのプログラムの推定結果の比較
- 注1: X軸が LOGIST による推定値  
Y軸が2分法による推定値をあらわす
- 注2:  $n=576$  向性に関するものだが、抑うつ性についても、ほぼ同様であった。

目), 因子負荷量の大きい項目が選ばれた。因子分析の結果, 第一固有値 (18.06) と第二固有値 (5.46) 以下の差が大きく, これらの項目の一次元性は一応保証されたと考えられた。

注3) ツングの質問紙の日本語訳は, 穂積 登監修「ウツのはじまり症候群」光書房, 1985 から採った。

注4) 語彙理解尺度のデータは東京大学の芝 祐順教授の御好意でお借りすることができました。なお, この尺度の被験者は中学1年生である。

## 文 献

- The American Psychiatric Association 1980 Quick reference to the diagnostic criteria from DSM III. APA. (高橋三郎ほか(訳) 精神障害の分類と診断の手引。医学書院, 1982)
- Comrey, A.L. 1957 A factor analysis of items on the MMPI depression scale. *Educational and Psychological Measurement*, 17, 575-585.
- Hambleton, R.K., & Swaminathan, H. 1985 *Item response theory: Principles and applications.*, Kluwer Nijhoff Publishing.
- 加茂富美子ほか 1965 MBTI 日本版標準化の試み(その1)(その2) 教育心理学会第7回総会発表論文集, 316-319.

Lord, F.M. 1977 Practical applications of item characteristic curve theory. *Journal of Educational Measurement*, 14, 117-138.

芝 祐順 1978 語彙理解尺度作成の試み。東京大学教育学部紀要第17巻, 47-58.

辻岡美延 1957 矢田部・Guilford 性格検査。心理学評論, 70-100.

Wood, R.L., Wingersky, M.S., & Lord, F.M. 1976 Research memorandum: LOGIST a computer program for estimating examinee ability and item characteristic curve parameters. Princeton, N.J.: Educational Testing Service.

Zung, W.W. 1965 A self-rating depression scale. *Archives of General Psychiatry*, 12, 63-70.

付記1 本研究では多くの被験者の方々と大学・高校の先生方の御協力をいただきました。貴重な時間をさいてくださった, 皆様に感謝いたします。

本研究は, 全面にわたって, 芝 祐順教授, 渡部洋助教授の御指導をいただきました。記して感謝の意を表します。また, 研究の過程でたえず助言を下された新潟大学竹綱誠一郎氏, 東京大学鎌原雅彦氏, 柴山直氏, 質問紙の作製を手伝って下さいました東京大学大学院の皆様にも感謝いたします。

付記2 本研究の計算には東京大学計算機センターを利用した。