

計量パーソナリティ心理学

第3章

GHQへの潜在ランク理論の適用

京都大学大学院教育学研究科 M2

西端和志

2017/05/17(水)

古典的テスト理論

- テキスト2章「項目反応理論による心理尺度作成」より

古典的テスト理論

古典的テスト理論で、今回重要になる用語

- 項目分析(Item analysis)

心理尺度・テストの各項目がどのような特徴を持っているか
調べるための統計アプローチ

主な指標に、項目困難度と項目識別力がある

- 項目困難度(Item difficulty)

各項目において「あてはまる」と回答した人の割合
(テストの場合は、問題ごとの正答率)

項目 j の困難度 p_j は

$$p_j = \frac{\text{「あてはまる」と回答した人数}}{\text{全回答者数}}$$

用語の確認

古典的テスト理論で、今回重要になる用語

- 項目識別力(Item discrimination)

その尺度によって測定しようとしている特性について、各項目がその特性の高い回答者と低い回答者をどの程度区別できるかを表す指標

識別力は r_j は、項目 j の反応パターンを u_j 、合計得点を x とすると、

$$r_j = u_j \text{ と } x \text{ の相関係数 (I - T相関)}$$

→1に近いほど、当該項目に「あてはまる」と答えた人の合計得点が高くなる傾向がある

用語の確認

- 困難度と識別力は全て手元のデータのみに基づいて計算されるので、いろいろ問題がある
 1. 別の集団でデータを取ると値は変わる(標本依存性)
 2. 尺度内の項目を一部変更してデータを取ると値は変わる(項目依存性)
- これらの依存性を克服する現代テスト理論として、項目反応理論や**潜在ランク理論**が提唱されている
 - 困難度と識別力を、モデルの母数として推定する

潜在ランク理論

尺度で人の心を評価する

- 心理尺度は、心の状態を潜在変数として測定する
 - 因子分析／項目反応理論などの統計モデル
- これらの統計モデルは、潜在得点は正規分布に従うという仮定を置いて、連続的な得点として潜在変数の推定を行う
- ただし、臨床場面では実用上、項目得点の合計値を因子得点とする簡易的な得点化を行うことが多い

臨床場面での二分法的スクリーニング

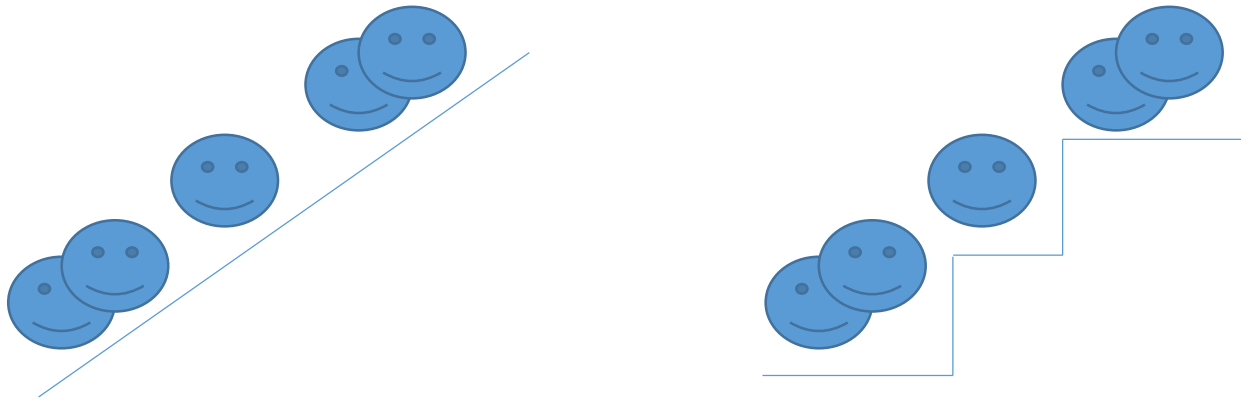
- 臨床場面ではしばしば、心理尺度を用いて二分法的スクリーニングを行う
 - 事前にカットオフポイントを定め、それに基づいて臨床群／非臨床群に分類
- しかし、固定されたカットオフポイントによるクライアントの二分化は問題がある
 - なぜ？
 1. 誤分類の可能性
 2. 適用場面における柔軟性のなさ

二分法 ⇒ 順序グループ(ランク)への分類

- その症状の程度に合わせて、クライアントを順序性を持ったいくつかのグループに分けることができれば、症状の理解や介入方法の工夫に繋がる

⇒**潜在ランク理論**(荘島, 2010)

- 潜在変数に順序グループ(ランク)を仮定し、二値や段階反応(e.g. リッカート尺度)のデータに対して、各回答者のそれぞれのランクへの所属確率を推定する手法
- 連続変量を2段階以上の順序尺度水準にできる



連続変量を順序尺度水準にすることの利点

1. ランクごとの非連続的な性質の記述ができる
 - 心理傾向の多くは連続性を仮定するが、実際にはある段階で質的に状態が変わることもある
2. 得点の“意味のある違い”が判断しやすくなる
 - e.g. 学力テストでA君は60点, B君は59点を取った
この時, A君の学力はB君より明らかに高いと言えるか
 - 順序尺度化すると, ランクごとの違いは明確になる
 - 得られるランクの意味を一定にすることで, 短縮版尺度とオリジナル版尺度における点数の解釈の違いも明確に

潜在ランク理論とその他の統計手法

- クラスター分析／潜在クラス分析との違い
 - これらは、参加者を順序性のない名義グループに分類する
 - 潜在ランク理論では、順序性を持ったグループに分類する
- 同じくテスト理論である、項目反応理論との違い
 - 項目反応理論は、症状に対して正規分布に従う連続変量を仮定
 - 潜在ランク理論では、症状に対して特定の分布を仮定しない

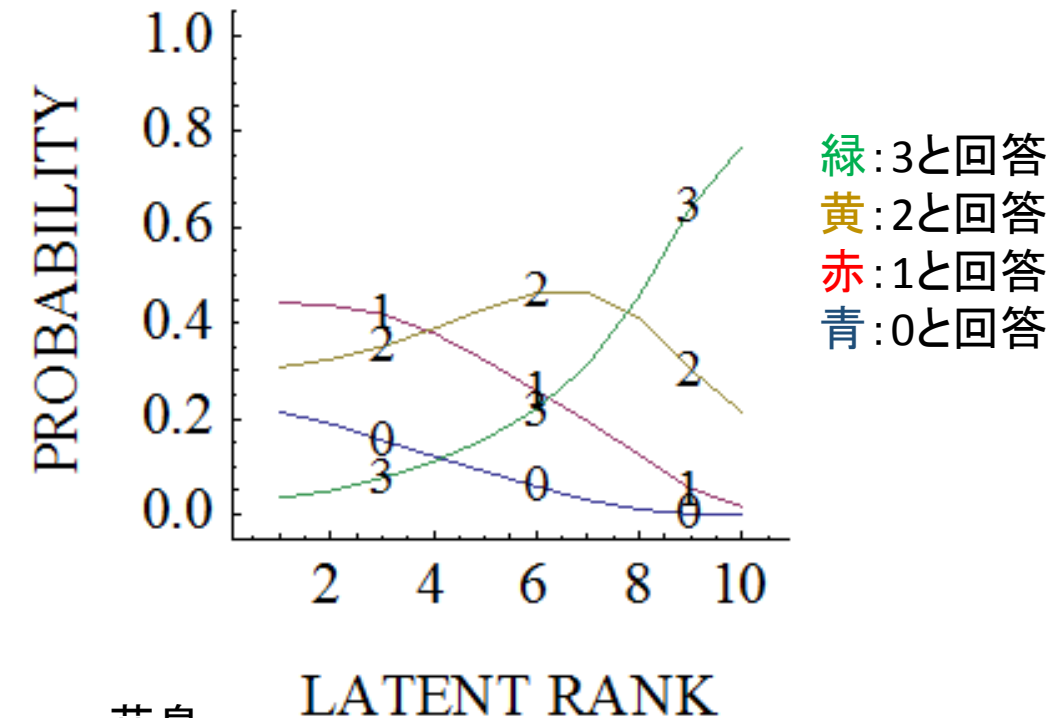
潜在ランク理論が推定するパラメータ

- 大きく3つ
 1. 項目の性質をあらわすパラメータ
 2. 尺度全体の性質をあらわすパラメータ
 3. 回答者の性質をあらわすパラメータ

項目の性質をあらわすパラメータ

項目カテゴリ参照プロファイル(ICRP)

- 各ランクに所属する回答者が、特定の項目の特定の反応段階に反応する確率



- 各曲線がICRP
- 縦軸は各反応段階への回答確率
- 横軸はランク

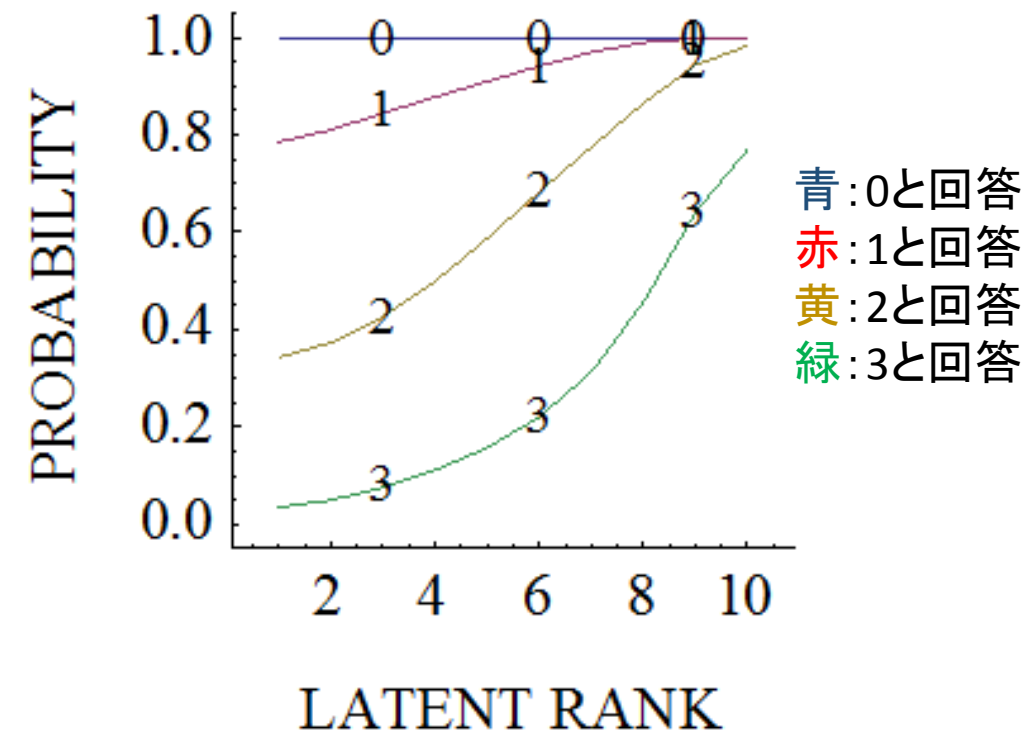
荘島

(<http://www.rd.dnc.ac.jp/~shoji/ma/ntt/introLRT-j.ppt>)

項目の性質をあらわすパラメータ

境界カテゴリー参照プロファイル(BCRP)

- 各ランクに所属する回答者が、各反応段階以上に反応する確率



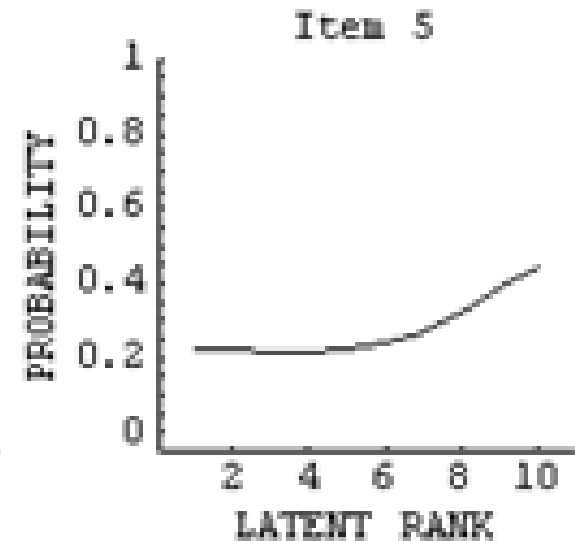
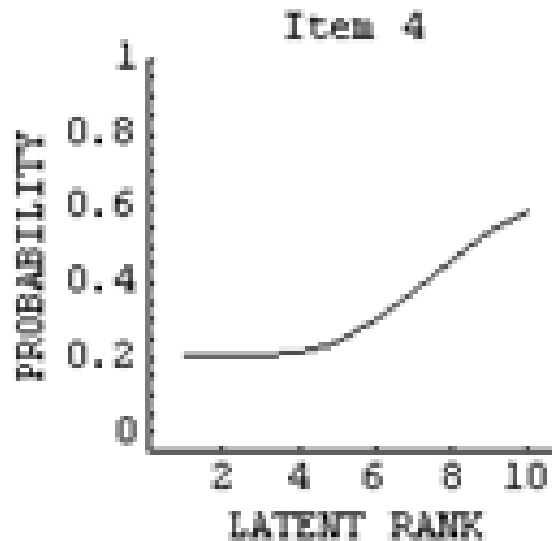
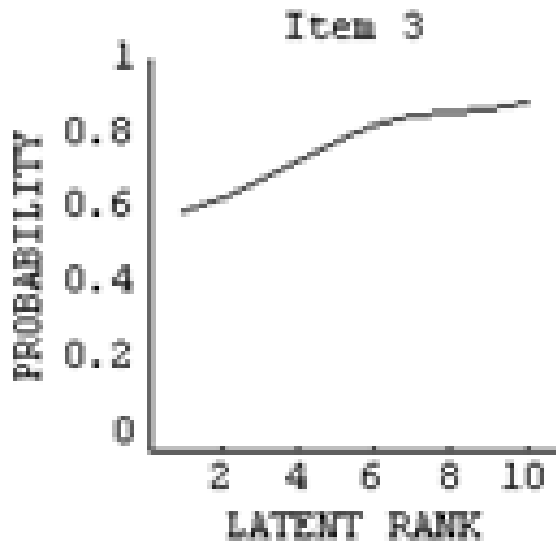
莊島

(<http://www.rd.dnc.ac.jp/~shojima/ntt/introLRT-j.ppt>)

項目の性質をあらわすパラメータ

項目参照プロファイル(IRP)

- ICRPに各反応段階の得点をかけて合計したもの
 - 各ランクの人が、各項目で平均的にどのぐらい得点をつけるかが分かる



莊島

(<http://www.rd.dnc.ac.jp/~shojima/ntt/introLRT-j.ppt>)

項目の性質をあらわすパラメータ

項目参照プロファイル(IRP)

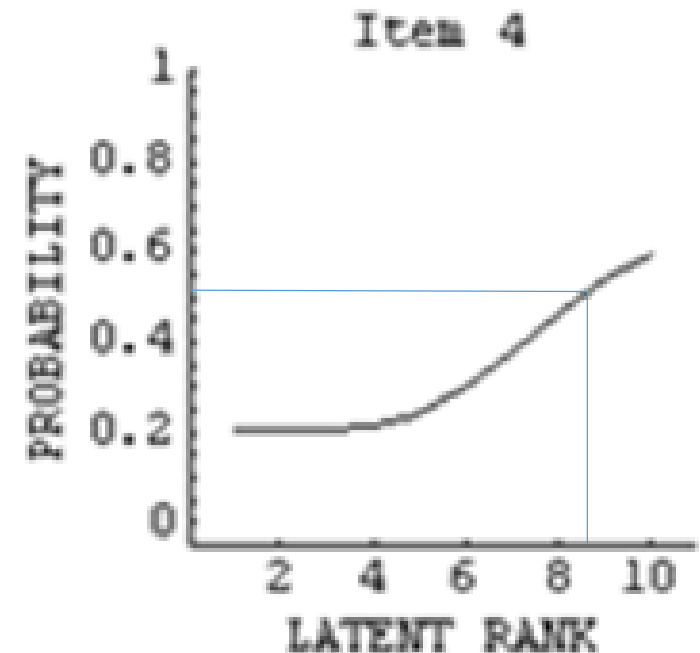
- IRPを用いて, 困難度と識別力を計算できる(これをIRP指標と呼ぶ)

IRP指標の困難度

- 評定得点の50%に最も近いランク
- 実情に合わせて, 評定得点の割合は適宜変更すると良い

IRP指標の識別力

- 隣合うランク間のIRPの変化が最も大きいランク(とその差そのもの)



莊島

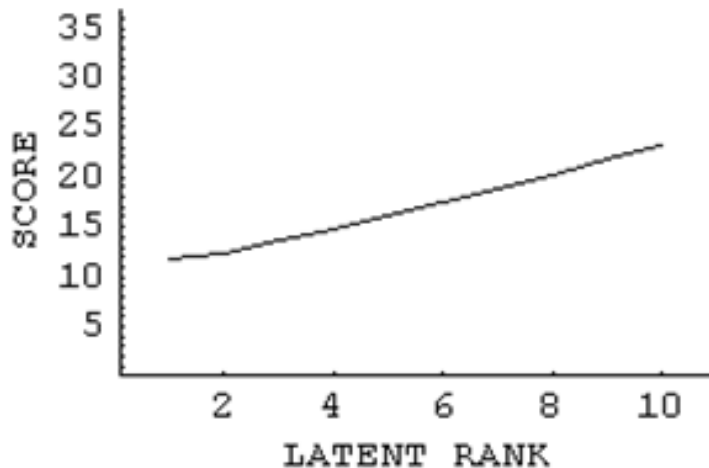
(<http://www.rd.dnc.ac.jp/~shojima/ntt/introLRT-j.ppt>)

※棒線は発表者

尺度全体の性質をあらわすパラメータ

テスト参照プロフィール(TRP)

- 各ランクに所属する回答者の、尺度合計得点の平均(各潜在ランクの尺度得点の期待値)



荘島

(<http://www.rd.dnc.ac.jp/~shojima/ntt/introLRT-j.ppt>)

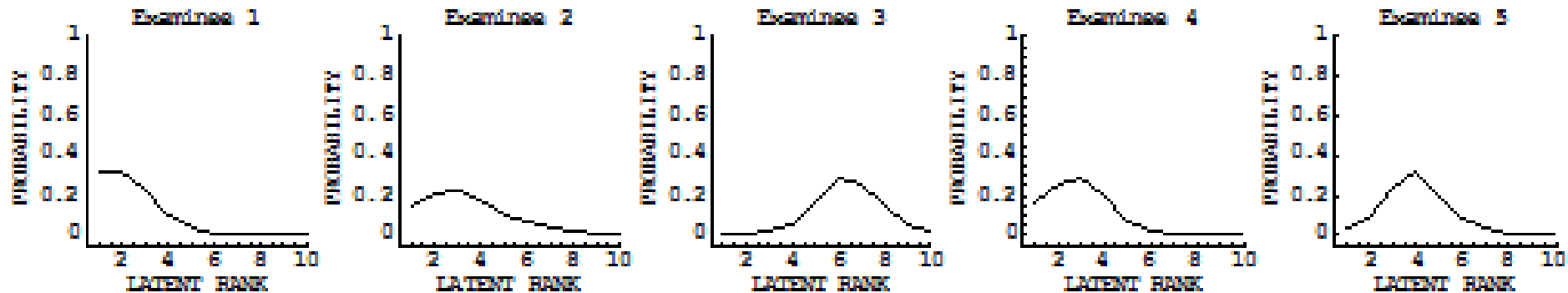
1. TRPが単調増加であることを, **弱順序配置条件**
2. TRPが単調増加かつ, 全項目のIRPも単調増加であることを, **強順序配置条件** と呼ぶ

弱順序配置条件が, 潜在ランク理論の必須条件

回答者の性質をあらわすパラメータ

ランク・メンバーシップ・プロファイル(RMP)

- 各回答者が, 各ランクに所属する確率



莊島

(<http://www.rd.dnc.ac.jp/~shojima/ntt/introLRT-j.ppt>)

- このように, 潜在ランク理論ではすべてのランクにおける回答者の所属確率を直接計算できる

潜在ランク理論の推定方法(荘島, 2010)

自己組織化マッピング(Self Organizing Mapping; SOM)

- 個別学習型のニューラルモデル
 - 計算の度に微妙に異なる結果を出力する
- 標本サイズが少ない時に使用する

生成位相マッピング(Generative Topographic Mapping; GTM)

- 一括学習型のニューラルモデル
 - 何回計算しても結果が同じ
- 計算が非常に早い
- 大規模データ向き

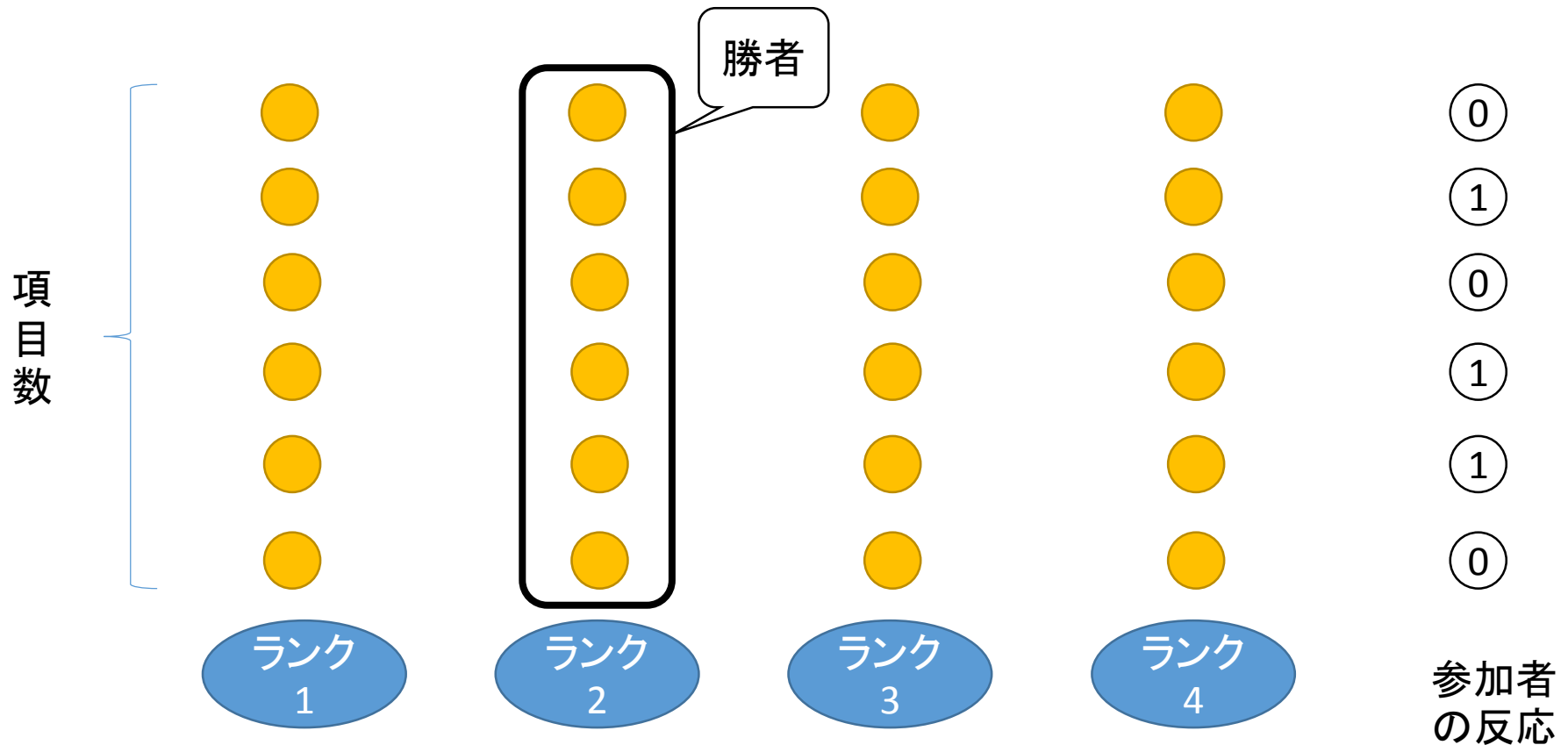
潜在ランク理論の模式図(SOMの場合)

1. 列数を潜在ランク数, 行数を尺度の項目数とする
ランク参照行列を作る



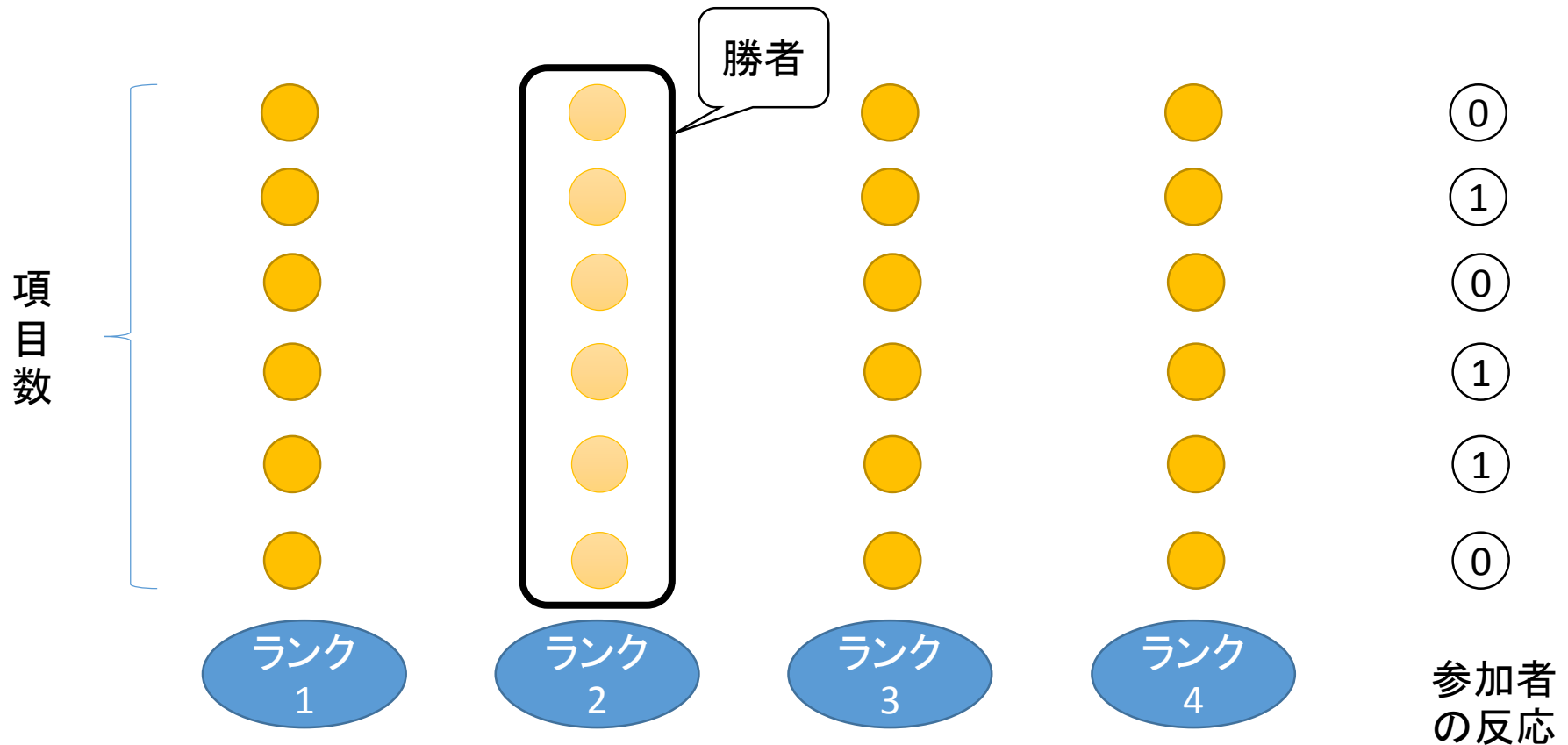
潜在ランク理論の模式図(SOMの場合)

2. 参加者の反応データを入力し, それと一番近いランク参照ベクトルを持つランク(勝者ランク)を決める



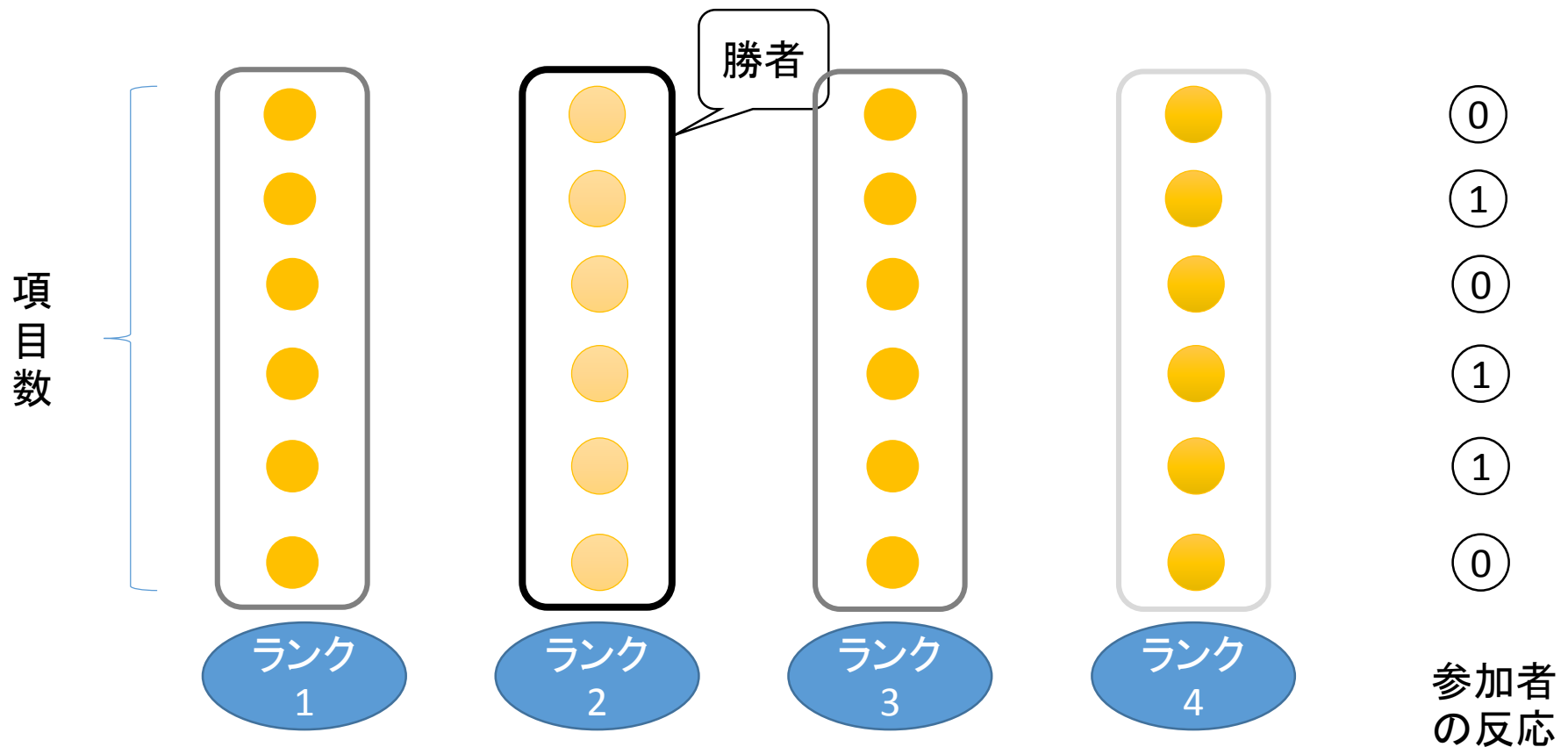
潜在ランク理論の模式図(SOMの場合)

3. 勝者ランクのランク参照ベクトルが，参加者のデータにより近くなるように更新する



潜在ランク理論の模式図(SOMの場合)

4. 勝者ランクに近いランクの参照ベクトルも，勝者ランクに近くなるように更新する



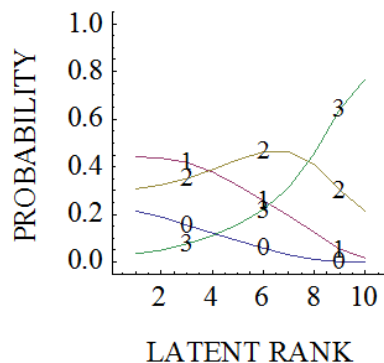
(1~4をデータの数だけ繰り返す)

潜在ランクの推定

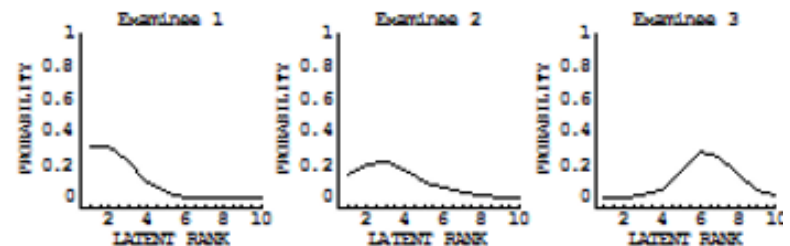
- 十分に大きな標本サイズ of 分析によって項目の特徴 (ICRP など) が推定されている場合, そのパラメータを用いて小さい標本における回答者の性質 (RMP) を推定できる

⇒ 極少数 (1 人から) のデータについても, 妥当な RMP を推定することができる

※ICRP(p.13)



※RMP(p.18)



精神的健康調査票(GHQ60)の分析

調査に用いる質問紙

精神的健康調査票(GHQ; 60項目)

- 精神疾患患者の評価に利用される質問紙
- 全体的なストレス反応や不安の程度を測定できる

使い方は2つ

1. スクリーニングテストとしてのGHQ

- 16-17点をカットオフポイントとし、健常群と臨床群を区別する
- 柔軟な運用が難しい

2. 連続変量としてのGHQ

- 0～60点の61段階でストレスを評価する
- 1点の差が意味のある差にならない

⇒順序尺度化すれば使いやすくなるのでは

方法

調査対象者

2012年実施調査

- マクロミルによるWeb調査(以降Web)：548名(男性276名, 女性272名, $M_{age}=42.99(SD=13.31)$)

1985年実施調査(中川・大坊, 1985)

- 健常者：55名(男性20名, 女性35名, $M_{age}=40.23(SD=12.34)$)
- 神経症：80名(男性36名, 女性44名, $M_{age}=35.01(SD=13.20)$)
- 大学生：266名(男性147名, 女性119名, $M_{age}=19.79(SD=1.24)$)

調査項目

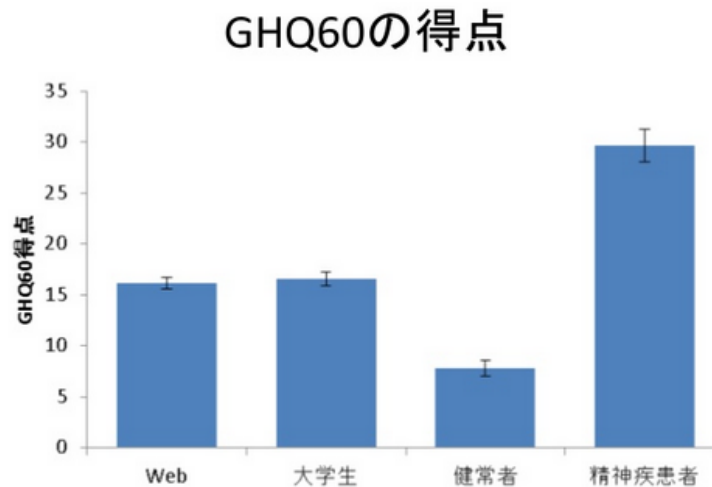
- GHQ60

- 最もストレスの程度が小さいものを0点, 最もストレスの程度が高いものを3点として測定(60項目4件法)
- 採点時は, $0 \cdot 1 \rightarrow 「0」$, $2 \cdot 3 \rightarrow 「1」$ として計算

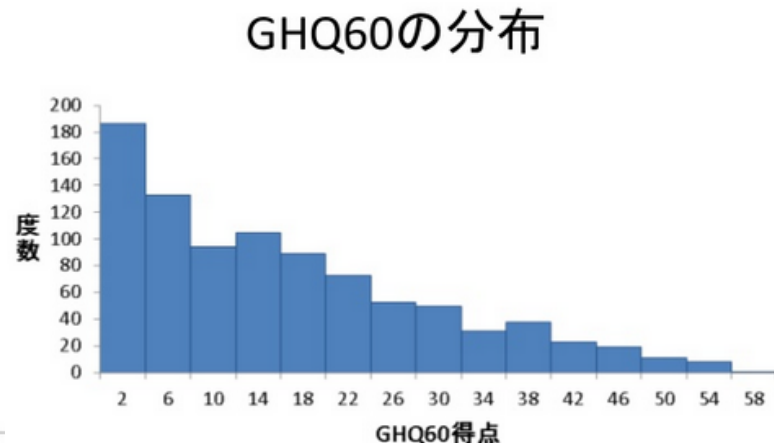
- 人生満足度

- 人生全体における満足度を測定(5項目7件法)

GHQ60のデータ要約



※エラーバーは標準誤差



コルモゴロフ・スミルノフ検定 ($D = 0.105, p < .001$)

潜在ランクの推定

潜在ランク数の判断

- 段階的潜在ランクモデルを実施
- 目標潜在ランク分布などの制約は設けず
 - ストレスの程度がどのような分布になっているかについて、明確な事前情報がないため
- 潜在ランクの弱順序配置条件が満たされていることを確認
- 潜在ランク数は、一番あてはまりの良い4つを選択
 - CAIC, BICなどを用いて

潜在ランクの推定

項目参照プロファイル(IRP)

IRP指標を以下のように設定

- 困難度：各項目について，参加者の平均的な反応が初めて1を超えるランク
- 識別力：各ランク間で最も大きいIRPの差
- 次のページから，算出デモ
……と行きたいところだったが，この解析で使用された生データが手に入らないので，別のサンプルデータでのデモを後で行う
ひとまず，この研究の話の続きを

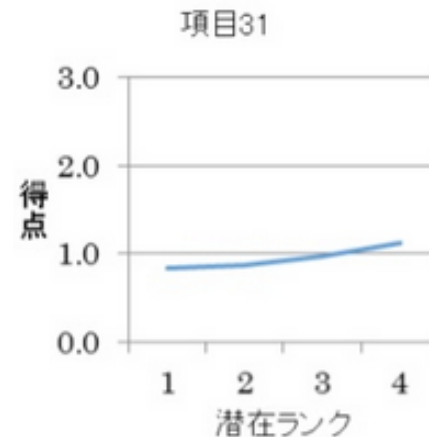
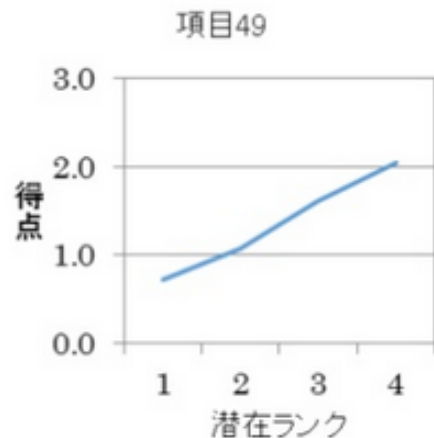
IRP指標の確認

- 識別力の高い項目

e.g. 項目49：「いつもより気が重くて、憂鬱になることは」

- 識別力の低い項目

e.g. 項目31：「いつもより周りの人に親しさや温かみを感じることは」



ランクごとの特徴の記述

- ランク1
 - 健常な人が所属するランク
- ランク2
 - 軽いストレス症状や社会活動障害が見られる
 - 弱い身体的症状
- ランク3
 - 不安や不眠
 - 強い身体的症状
- ランク4
 - うつ症状



清水(2014) (<http://norimune.net/2199>)

ランクメンバーシップと回答者の所属ランク

- 参加者タイプ別の各ランクの度数

Residual analysis of GHQ rank

Rank	Participant type				Overall
	Web	College students	Healthy person	Neurotic patients	
1	△ 187 (62.75)	75 (25.17)	△ 35 (11.74)	▼ 1 (0.34)	298
2	144 (54.75)	85 (32.32)	16 (6.08)	18 (6.84)	263
3	103 (56.28)	54 (29.51)	▼ 4 (2.19)	22 (12.02)	183
4	114 (55.61)	52 (25.37)	▼ 0 (0.00)	△ 39 (19.02)	205
Overall	548	266	55	80	949

清水・大坊(2014)

- 括弧内の数値は、各セルの人数を各ランクの総人数で割ったもの
- △は期待度数に比べて、観測度数が有意に大きかったセル
- ▼は期待度数に比べて、観測度数が有意に小さかったセル

ランク間のGHQ得点の差の検定

Table 3
Means and standard deviations of GHQ scores for each rank

Rank	N	M	SD	95%CI Lower	95%CI Upper	Threshold
1	283	4.61	3.30	4.01	5.21	9/10
2	255	11.21	4.94	10.58	11.84	16/17
3	179	21.81	3.83	21.06	22.56	28/29
4	197	37.34	7.85	36.62	38.06	—

清水・大坊(2014)

- 隣接するランク間の差は、全て1%水準で有意
- ※効果量(論文のデータを元に発表者が計算)

	効果量 - g	95%CI
ランク1-2の差	1.58	1.39-1.78
ランク2-3の差	2.34	2.10-2.59
ランク3-4の差	2.47	2.20-2.74

ランク間のGHQ得点の差の検定

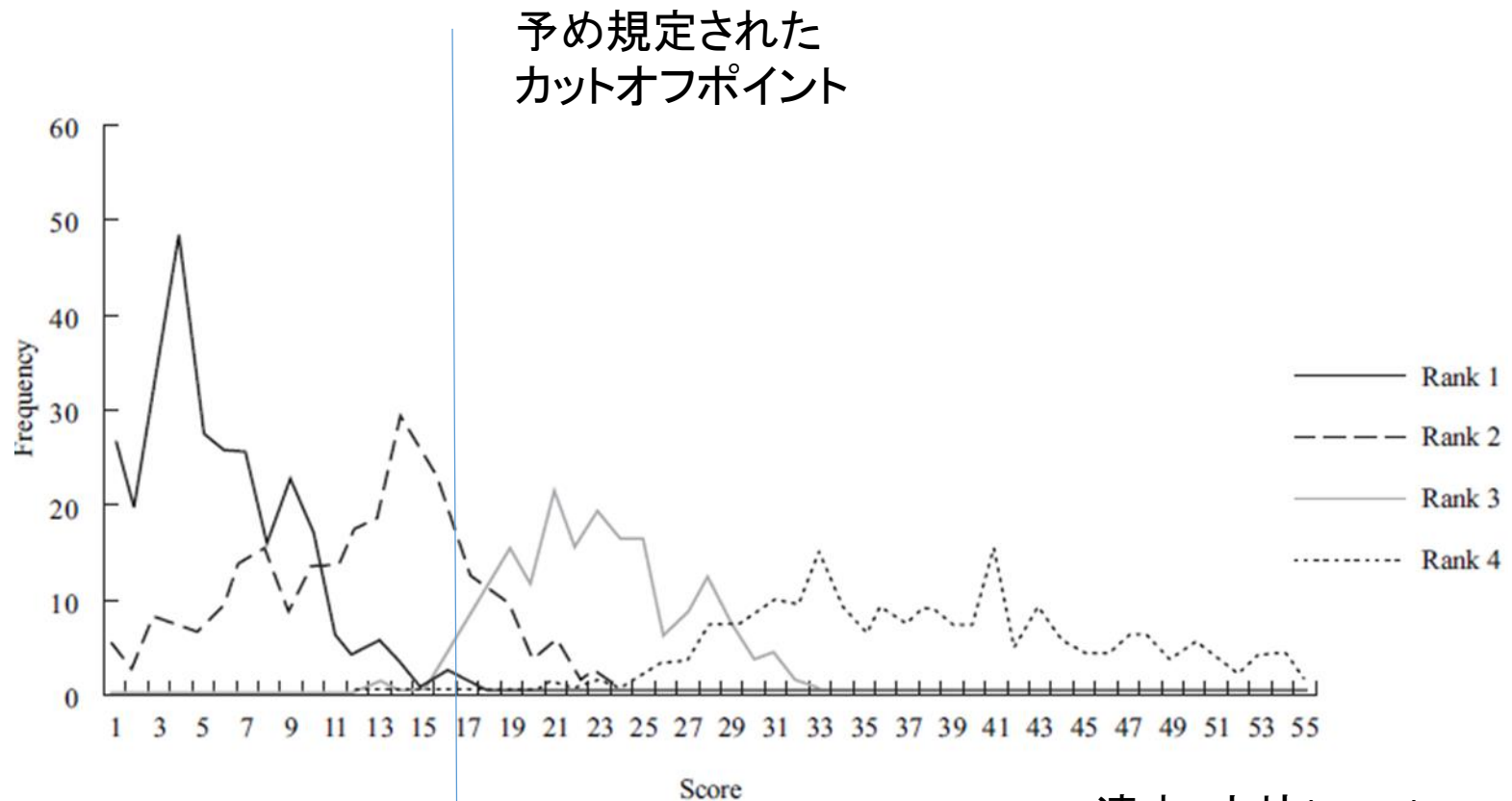
- 効果量(清水・大坊(2014)で報告されたもの)

	効果量 - g	95%CI
ランク1-2の差	1.29	1.10-1.47
ランク1-3の差	3.35	3.09-3.61
ランク1-4の差	6.38	5.96-6.80

- ランクには得点に順序性があり, かつその差には実質科学的な意味があると考えられる

ランクメンバーシップと回答者の所属ランク

ランク別のヒストグラム



清水・大坊(2014)
※棒線は発表者

人生満足度や疾患の有無との順位相関

- 人生満足度-GHQ60 : $r = -.43$, 95%CI $[-.46, -.39]$
- 人生満足度-潜在ランク : $r = -.42$, 95%CI $[-.45, -.38]$
- 精神疾患の有無-GHQ60 : $r = .25$, 95%CI $[.22, .28]$
- 精神疾患の有無-潜在ランク : $r = .25$, 95%CI $[.22, .28]$

短縮版による潜在ランクの推定

1. GHQ60のIRPをそのまま短縮版のIRPとして用いて,
潜在ランクを推定
 - GHQ28の推定の適合度はRMSEA=.049
 - GHQ12の推定の適合度はRMSEA=.064

2. GHQ60と各短縮版との, 潜在ランクのスピアマンの
順位相関係数を算出
 - GHQ60-GHQ28 : $r=.94$, 95%CI[.93,.94]
 - GHQ60-GHQ12 : $r=.91$, 95%CI[.90,.92]

潜在ランク理論の評価まとめ

二分法と比較して

- ランクごとの柔軟な臨床介入が可能に
- ランク数の決定を，適合度指標などの客観性のある指標を用いて吟味できる
- ランクごとの特徴の記述により，クライアント自身が自分の症状を理解することが容易に

連続得点評価と比較して

- 尺度水準を順序尺度に落としても，人生満足度などの予測力はほとんど変わらず
- 短縮版を利用する際も，元尺度のICRPなどを利用することで，オリジナル版と同様のランクの解釈を行える

潜在ランク理論を実際に使ってみる

実習：潜在ランクの推定

必要なアプリのダウンロード

- Exametrika5.3(荘島, 2014)
 - 大学入試センターの荘島宏二郎先生が作成したアプリ
 - 以下のURLに行き, ダウンロードをクリック
 - <http://antlers.rd.dnc.ac.jp/~shojima/exmk/jindex.htm>

エグザメトリカ

exametrika

English Site

トップ

潜在ランク理論

項目反応理論

非対称三角尺度法

カテゴリカルデータ解析

ダウンロード

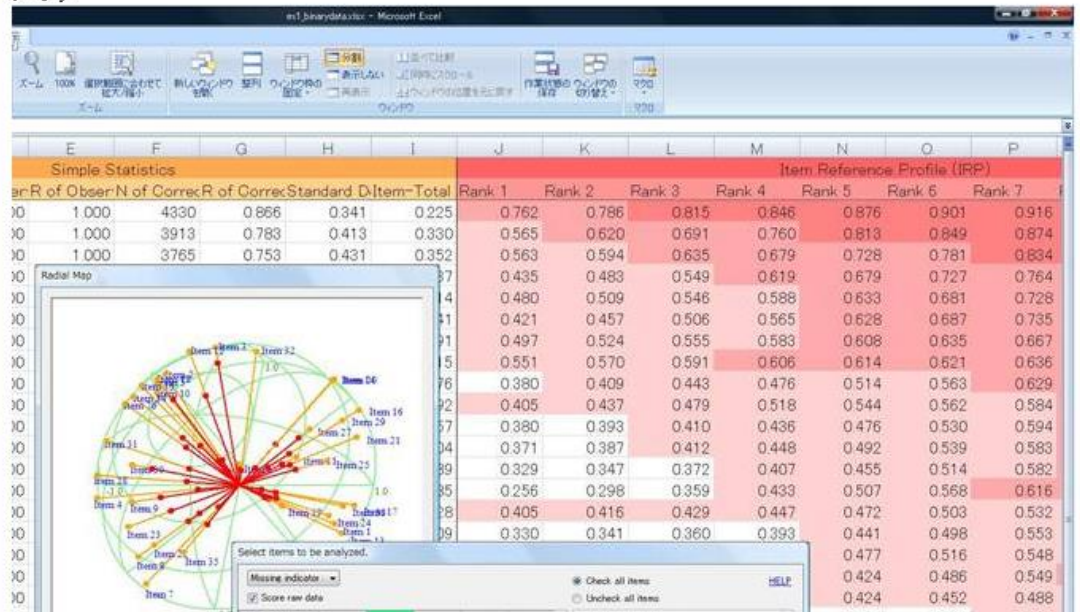
ヘルプ

リンク

shojima@rd.dnc.ac.jp

エグザメトリカ

エグザメトリカは、潜在ランク理論(ニューラルテスト理論)や項目反応理論、非対称三角尺度法のためのソフトウェアです。



潜在ランクの推定

必要なアプリのダウンロード

- Exametrika
 - 最新版をダウンロード

エグザメトリカ

exametrika

[English Site](#)

[トップ](#)

[潜在ランク理論](#)

ダウンロード

エグザメトリカを実行するには、Microsoft Windows Vista以降が必要です。また、Microsoft .Net Framework 4以降が必要です。 .Net Frameworkはマイクロソフト・ダウンロードセンターから入手できます。

データは、Microsoft Excelで用意してください。ただしExcelは2007以降です（拡張子が.xlsではなく.xlsxです）。

バグを発見したら、ぜひ教えてください。以下のプログラムは、エグザメトリカのソースです。ご自由に使ったり、書き換えたりして下さってけっこうです。もしも、これらのソースを利用してプログラムを開発したときは、エグザメトリカを引用してください。また、プログラム中、ご不明な点があれば、お問い合わせください。

[LRT-GTM](#) (エクセル用5.3版)

[LRT-SOM](#) (エクセル用5.3版)

更新日	版	Excel (xlsx)	ノート
2014/09/30	5.3	DL	LRT分析で事前分布を指定することが可能になりました。
2011/05/27	5.2	DL	ATRISCAL分析の推定方法を修正しました。 .Net Framework 4以降が必要です。
2011/04/24	5.1	DL	カテゴリカルデータ解析の分析機能を追加しました。

潜在ランクの推定

- Exametrika(exmkj53.exe)を起動
- Exametrikaに同封されているサンプルデータ(ex3_gradeddata.xlsx)をドラッグして、下図の青くなっているところにドロップ
- 「データ・ワークシート」をクリックし、「graded data」にする(データがあるシートを選択する)



潜在ランクの推定

- 左上の「欠測指示子」を「(dot)」に変更
- データの起点として、下の矢印の部分をクリック
- 「セット」を押す⇒右にデータが出るので、右の「セット」も押す

データの起点を指定してください

(dot)

☐ 生データを採点する

☒ 全項目をチェックする [HELP](#)

☐ 全てのチェックを外す

	Item 1	Item 2	Item 3	Item 4
1	2	2	3	0
2	2	1	2	.
3	3	2	2	2
4	2	0	3	.
5	2	2	1	.
6	1	2	0	0
7	2	2	3	2
8	1	2	2	1
9	2	3	3	.
10	1	1	2	.
11	2	3	1	.
12	2	1	2	0

ラベル N K C1

戻る セット セット

潜在ランクの推定

LRT-SOMタブを選択し、以下のように設定

- 推定の設定で、段階モデルを選択
- 潜在ランク数を(ひとまず)10に指定
 - 潜在ランク数を減らすと弱順序配置条件を満たしやすくなる
- 事前分布について、「一様分布」を「指定しない」に変更
- 出力オプションを全部クリックする
 - (上記要素の詳しい説明はExametrikaのWebページに↓
<http://antlers.rd.dnc.ac.jp/~shojima/exmk/jindex.htm>)

分析実施 ⇒ Excelファイルができる

潜在ランクの推定

- できたExcelファイルを開く
- 「Summary」のシートを見ると、弱順序配置条件や強順序配置条件が満たされているかが分かる
- 「Test」のシートを見ると、モデルの適合度が確認できる
 - CAICなどを頼りに、より適切な潜在ランク数を探せる
- 「Item」のシートを見ると、各項目のIRPとIRP指標が確認できる
- 「Examinee」のシートを見ると、参加者ごとのRMPと、所属していると推定される潜在ランクが分かる

おまけ Rで潜在ランクの分析

- 今回は紹介できなかったが、Rで潜在ランクの分析をすることも可能
- 詳細は清水(2015)
<https://www.slideshare.net/simizu706/latent-rank-44708240>
- あるいは木村・登藤・荘島(2015)
<http://www2.kansai-u.ac.jp/jart2015/files2015/ippan/A0054.pdf>
- ExametrikaはWindows環境を要求するので、非Windows派は是非

参考文献

久保沙織(2017). 項目反応理論による心理尺度の作成 荘島宏二郎(編) 計量パーソナリティ心理学(pp.19-43.) ナカニシヤ書店

清水裕士・大坊郁夫(2014). 潜在ランク理論による精神的健康調査票(GHQ)の順序的評価 心理学研究, 85, 464-473.

荘島宏二郎(2010). ニューラルテスト理論 植野真臣・荘島宏二郎(編) 学習評価の新潮流(pp.83-111.) 朝倉書店

荘島宏二郎(2014). Exametrika5.3
(<http://antlers.rd.dnc.ac.jp/~shojima/exmk/jindex.htm>; 2017年5月12日閲覧)

加えて, 清水裕士先生のWebページ(<http://norimune.net> ; 2017年5月12日閲覧),

荘島宏二郎先生のWebページ
(<http://www.rd.dnc.ac.jp/~shojima/ntt/jindex.htm>; 2017年5月12日閲覧)

を随時参考にしました