研究委員会企画チュートリアルセミナー

項目反応理論（IRT）の考え方と実践

―測定の質の高いテストや尺度を作成するための技術―

企画・司会：宇佐美慧（東京大学）
企画：荘島宏二郎（独立行政法人大学入試センター）
話題提供：光永悠彦（名古屋大学）
話題提供：登藤直弥（筑波大学）

お願い

- 本チュートリアルの後半では統計ソフト R 内のパッケージを用いて、データ分析を実際に行います。そこで使用する、

psych　irtoys　ltm　plink

の 4 つのパッケージについて事前にインストールを済ませておいてください（インストールについては補足資料もあります）。また、分析で使用するデータについても、

http://satoshiusami.com/Data.zip

から

bfi.csv　LSAT.csv　PARA.csv　登藤発表資料.pptx

のファイルが入手できますので予めダウンロードしてください。本チュートリアルでは R の操作に関する基礎的な説明も行いますが、ごく初歩的な操作経験は既にあることを前提と致します。

はじめに

- 項目反応（応答）理論（item response theory：IRT）…資格試験・能力試験・臨床検査・学力調査といった国内外の様々なテストの開発・評価に活用されている方法論。

- 特に、受験者数が数千・数万以上と大規模なテストの継続的な運用を行う上でのグローバル・スタンダード。

- IRT はLord(1952)によってその基本的な枠組みが整備された。IRTは、計量心理学（psychometrics）を軸にして体系化されてきたテスト理論（test theory）の一領域として発展を重ね、国内外のテスト開発の中に徐々に普及していった。

IRTのテストへの活用例

- 英語の民間試験（TOEIC・TOEFL・GTEC・CASEC）
- 外国語としての日本語に関する公的試験（日本語能力試験・BJTビジネス日本語能力テスト）
- 学力調査（PISA・TIMSS・NAEP）
- その他、日本留学試験・ITパスポート試験など
- テストだけではなく、心理尺度・心理検査の作成・評価にも広く活用されている。

IRTへの関心の高まりとソフトウェア

- 大学入学共通テストなど、新しい高大接続のあり方をめぐる議論の中でも、IRT やコンピュータ・ベースト・テスティング（CBT）のような、テスト理論に関わる用語が頻繁に登場し、社会的にも広く認知されるようになってきた。

- IRT に関する日本語の優れた教科書・解説書も既に多く出版されており、さらにはBILOG-MGやIRTPROなどのIRTに特化した商用ソフトや、フリーの統計ソフトRにおけるltmパッケージなど高度なデータ分析環境も整備されつつある。

IRTに関する最近の教科書

- 加藤健太郎・山田剛史・川端一光 （2014）．Rによる項　目反応理論　オーム社
- 光永悠彦 （2017）．テストは何を測るのか：項目反応理論の考え方　ナカニシヤ出版
- 村木英治 （2011）．項目反応理論 朝倉書店
- 野口裕之・渡辺直登（編著）（2015）．組織・心理テスティングの科学：項目反応理論による組織行動の探求　白桃書房
- 豊田秀樹 （2012）．項目反応理論―入門編―（第 2 版） 朝倉書店

解説資料も多くあります。最近のものでは、例えば、

- 宇佐美慧 (2018). 潜在的な特性をモデルに組み込む　―項目反応理論の役割と近年の展開― 統計, 8, 16-22.

IRTは何がいいのか？
―テストを運用する際の問題―

- 全国模試が実施され、ある学校の平均点が、80点であった昨年度よりも伸びて、今年度は90点であった。その学校の生徒達の学力が平均的に高くなったと考えられるか？→×(テストが簡単になっただけかも)

- 仮に難易度が同じ２つのテストがあったとして、ある受験者が一方のテストを夏に受験して80点、また他方のテストを秋に受験して90点であったとする。では、受験者集団の中におけるこの受験者の順位（偏差値）も上昇しているか？→×（他の受験者の方が平均的により高く得点が上昇しているかも）

- テスト得点(例えば90点)が与える情報は、当該のテスト問題や受験者集団の性質に依存したものとなる。

ならば…

- 「８割以上の得点をとれば合格」・「上位２割以内に入った者を合格とする」といった基準は不適切？

- IRTでは、テスト（尺度）を受けた集団や問題（項目）に直接的に依存しない評価が可能。

→例えば、「80点」の意味は常に同じになる。

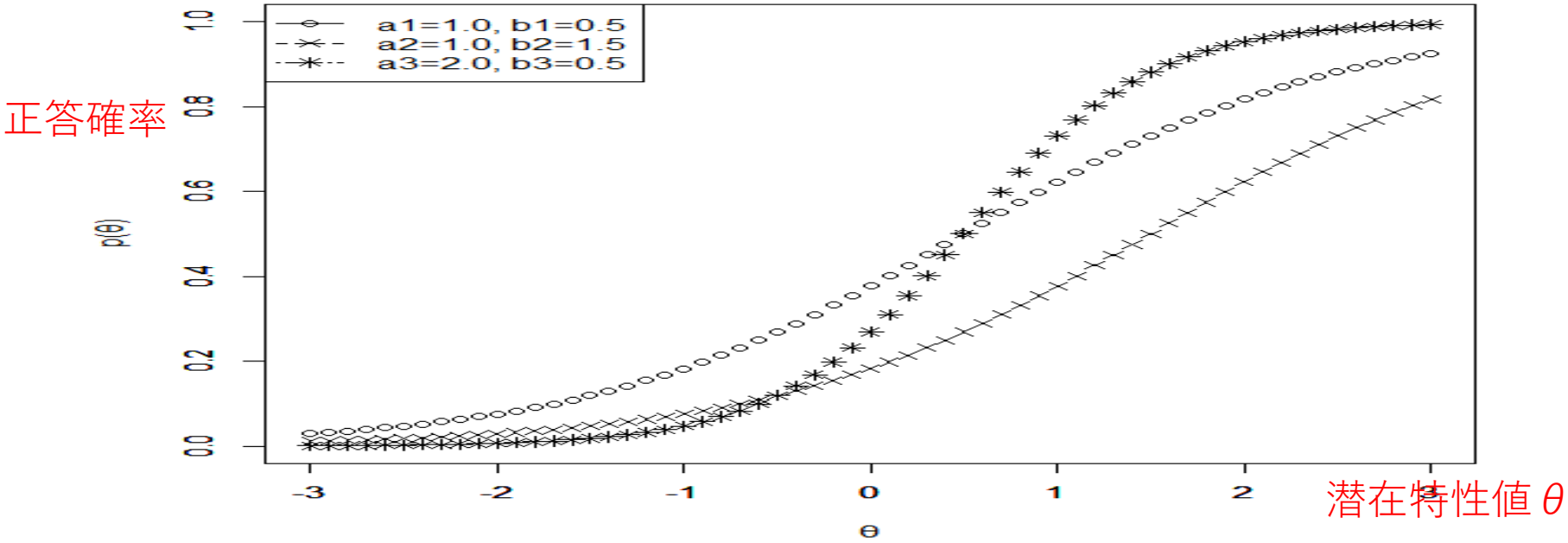
→回答した項目が互いに同一でなくても、集団間の能力差や、集団・個人の時間的な能力変化の評価が可能。

→大規模テストや学力調査への活用。

- このような一見不可能なことを実現するための核となるのが、項目反応モデルと呼ばれる数理モデルと、等化と呼ばれる手続きである。→本日のメインテーマ

項目反応モデル

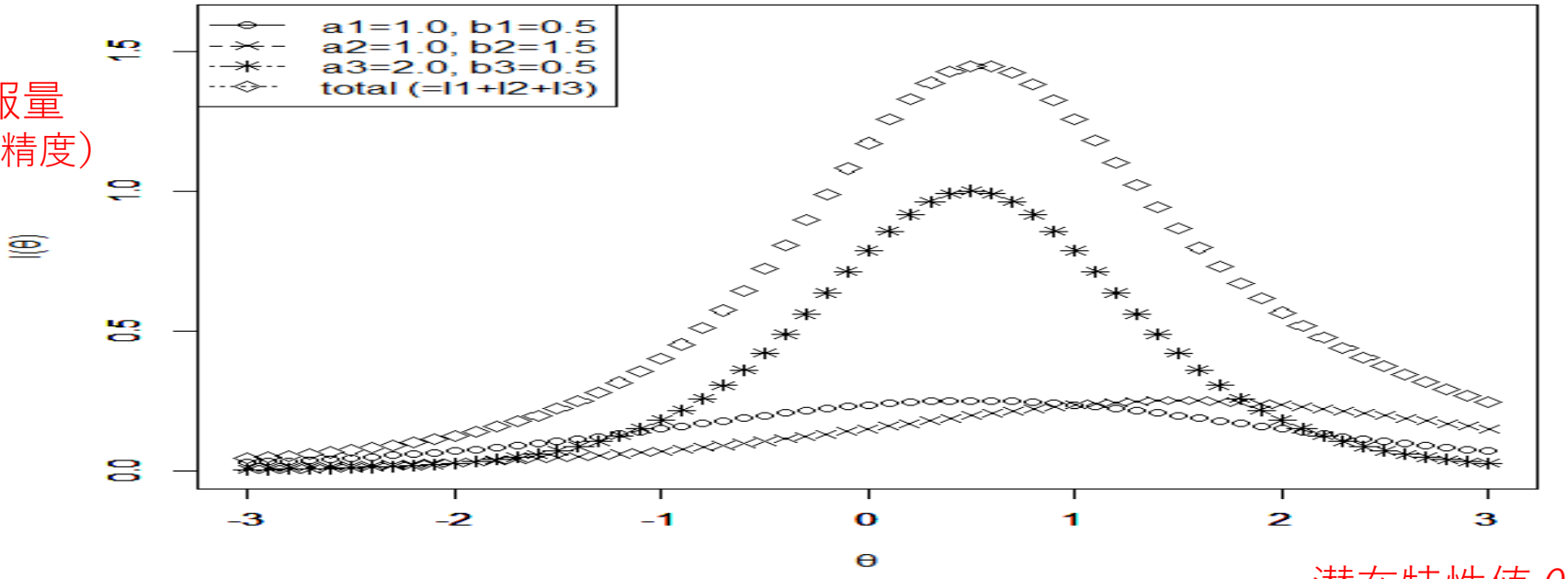
- 項目反応モデル…個人の能力・適性・態度等の直接観測できない潜在特性値をモデルに組み込み、ある潜在特性値 θ を示す個人がある項目（問題）に正答する確率を表現。



- IRT は、因子分析モデルのように潜在変数を伴うモデル。
- モデル化するということ→問題の難易度などの統計的性質を、受験者集団に依存しない数値情報として表現！

情報関数による測定精度（標準誤差）の評価

- 従来の評価方法では、「このテストには±10点ほどの誤差がある」・「テストの信頼性は0.7である」といった、全体的な意味での評価しかできない。
- IRTでは情報関数（後述）により、「能力が高い群では±３点程度の誤差が、また能力が低い群では±８点程度の誤差がある」と言ったような個別的な精度評価が可能。



情報量
（測定精度）

潜在特性値 θ

IRTのその他の利点

- テストの測定精度の情報を、受験者の能力（潜在特性値）の水準別に評価できる。

→テストに有用な項目の診断・取捨選択をより合理的に。

→（精度を極力担保しながら）テストの短縮版の作成。

→（同一の構成概念の測定を意図する）複数のテスト（尺度）の測定論的性質の比較が多角的にできる。

→大規模テストはもちろん、心理尺度・心理検査の作成・評価においても強みとなる。

→テストの難易度や得点分布を事前にコントロール可能。

→測定の効率化の観点から、各受験者の能力・回答状況に応じた項目の出題が可能（適応型テスト:CAT)。

→測定の妥当性向上や反応プロセスの評価、集団間比較等を目的とした統計（測定）モデルとしての活用。

企画趣旨

- 益々のIRTの隆盛が感じられる昨今ではあるが、国内における計量心理学の専門家の数は十分ではない（宇佐美, 2016 教心年報）。

- 幸いテスト理論の専門家は日本に数多くいるが、それでもテスト理論は計量心理学の一領域であり、大学・大学院教育の中でもテスト理論やIRTに十分な時間が割けられないのが実態であろう。

- IRTを用いたテストや尺度の開発・運用の実際に関しては、例えばこれらに関心のある実務家や他分野の研究者らの間であっても、まだ十分に共有されているとは言えない。

- 本チュートリアルではIRTを初めて学ぶ方を対象に，IRTの基礎と利点・注意点とともに，実際の心理学研究や実践例を基にしたRの実行方法についてなるべく網羅的かつ平易に解説する。

→光永悠彦先生…IRTの基本的事項と等化について(70分ほど)

→登藤直弥先生…Rを使ったIRTの実行について(50分ほど)

→質問時間(10-15分ほど)

日本教育心理学会　第60回大会
研究委員会企画　チュートリアルセミナー

「項目反応理論（IRT）の考え方と実践」
―測定の質の高いテストや尺度を作成するための技術―

項目反応理論の考え方

光永　悠彦
（名古屋大学　大学院教育発達科学研究科）

2018/09/15
慶應義塾大学　日吉キャンパス

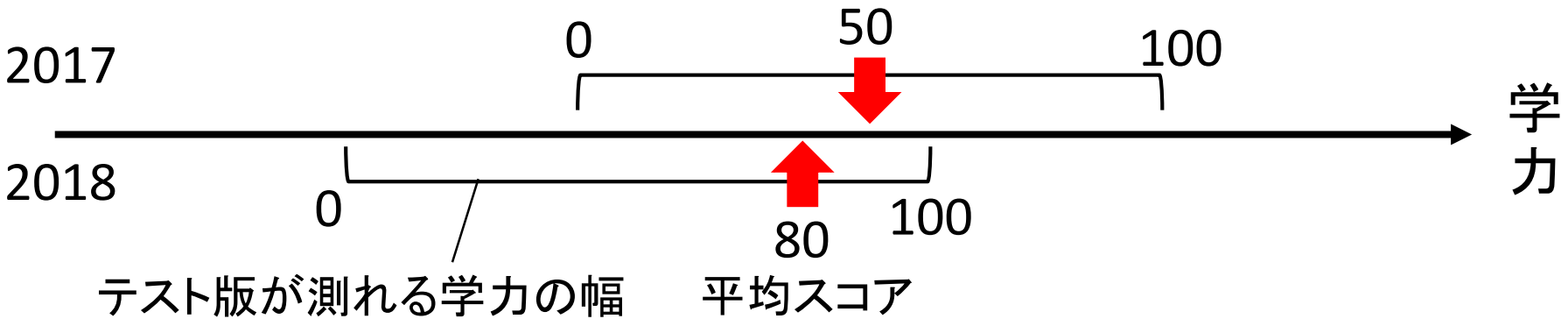
本日の内容

- 項目反応理論(IRT)を用いる理由
- IRTの概要
  - IRTによる尺度化
  - 母数の推定方法
  - 母数の推定精度とテスト情報関数
  - 実践例
  - 項目反応モデルの拡張
- IRTによる等化の概要
  - 等化の必要性
  - 等化の方法

- 項目反応理論(IRT)を用いる理由
- IRTの概要
  - IRTによる尺度化
  - 母数の推定方法
  - 母数の推定精度とテスト情報関数
  - 実践例
  - 項目反応モデルの拡張
- IRTによる等化の概要
  - 等化の必要性
  - 等化の方法

学力調査

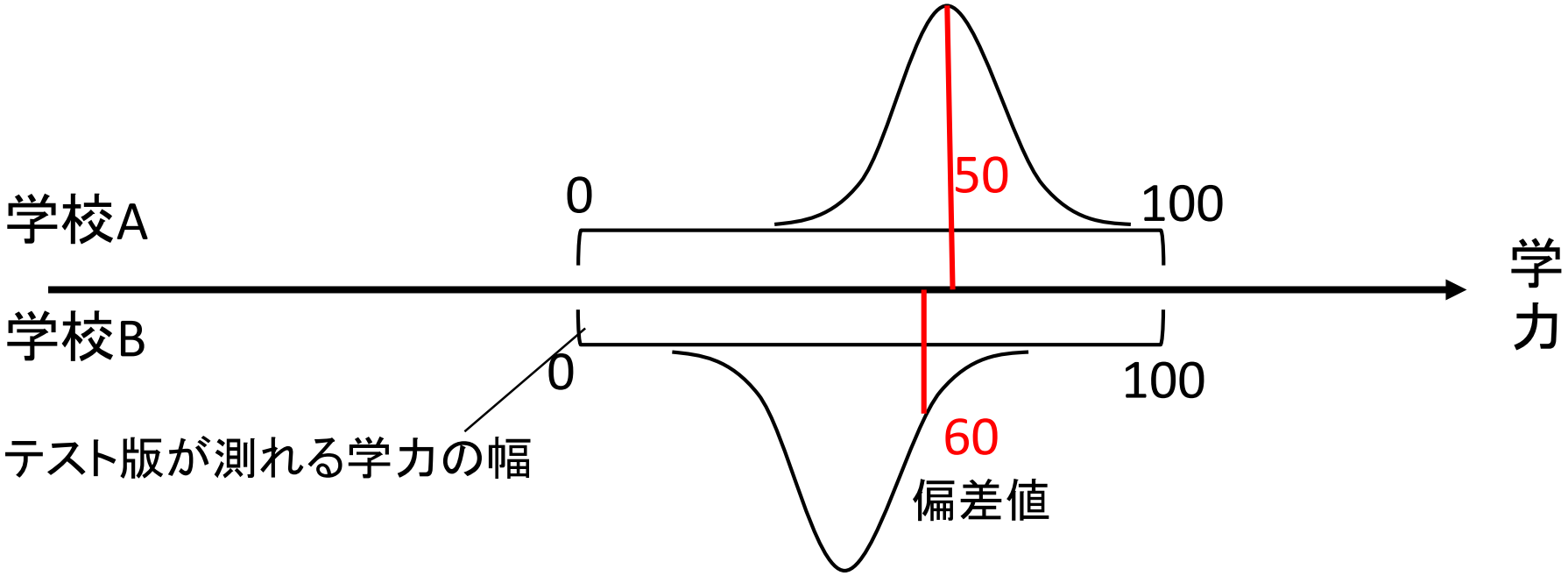
- 2017年度学力調査の結果，学校A全体の平均スコアが50点
- 2018年度学力調査の結果，学校A全体の平均スコアが80点
- 毎年の学力調査は，異なる問題項目群（テスト版）を用いて行われている
- 学校A全体において，学力が向上していると言えるか？



項目の困難度が調査実施年度によって異なる場合，
平均スコアを直接比較できない。

困難度を統制した場合

- ある県の学力調査は，困難度が同じになるように調整されている
- 学校Aにおけるxさんの偏差値が50
- 学校Bにおけるyさんの偏差値が60
- yさんの学力が，xさんよりも上であると言えるか？



素点を偏差値に換算したとしても，
学校ごとの学力分布が異なる場合，直接比較できない

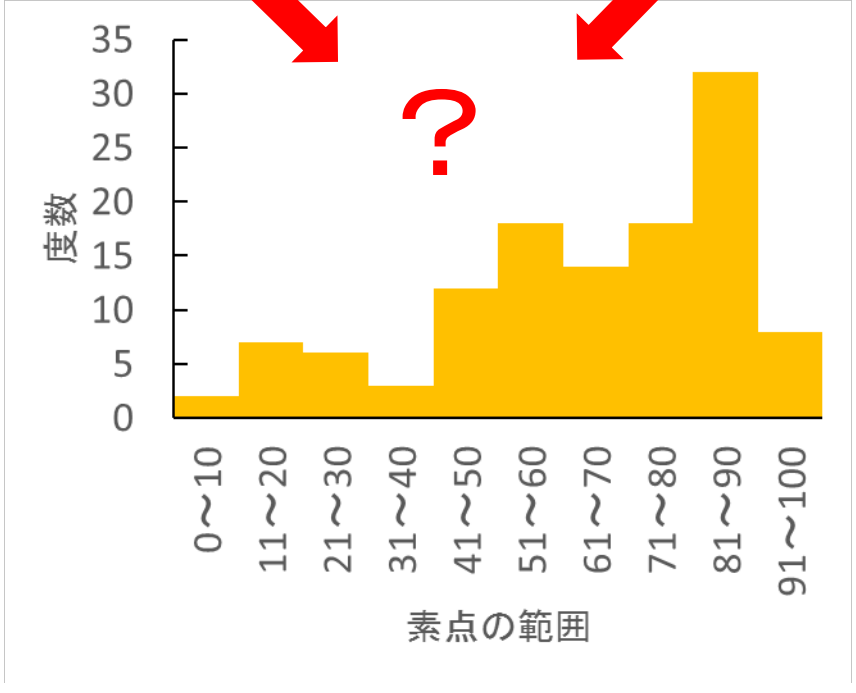
受験者能力の分布によらない項目特性

《古典的テスト理論》

高い素点の平均が，どちらの理由で得られたのかがわからない

低い
項目困難度

高い
受験者能力

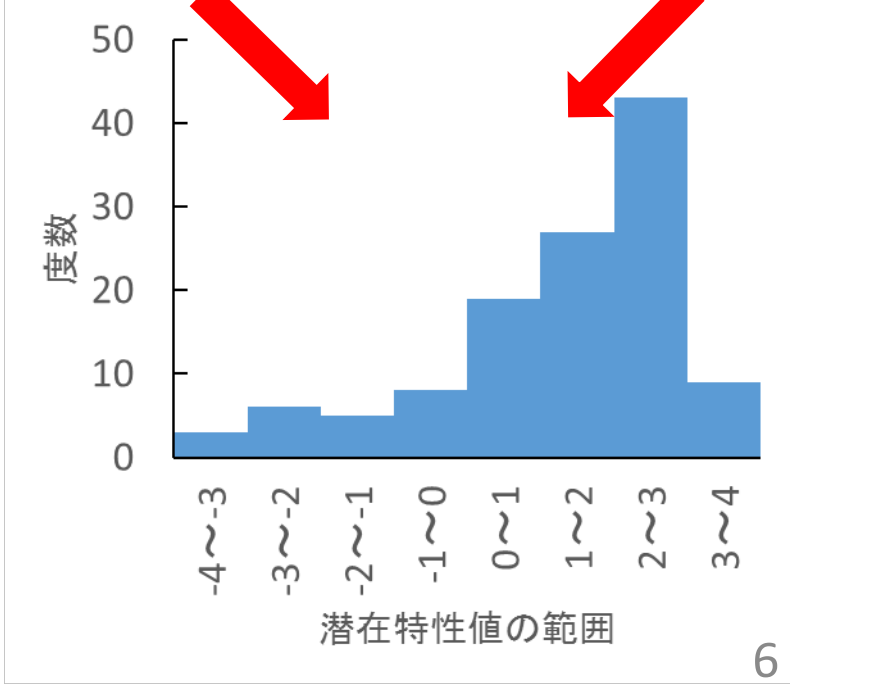


《項目反応理論》

項目反応モデル（後述）に基づき，受験者の能力の分布によらない項目特性を推定する

項目困難度
平均0.5

受験者能力
平均2.8



スコアの表示方法と解釈可能性

《古典的テスト理論》

- 素点（配点つき合計点）
  - 項目の困難度に依存
- 偏差値
  - 受験者集団全体の能力分布に依存

《項目反応理論》

- 受験者集団の能力分布に依存しないスコア（潜在特性値 θ）
- 出題した項目の難しさや，識別力（後述）などを「項目特性」（項目母数）として定義しておき，これらを根拠にして潜在特性値を推定する
- 異なるテスト版をまたいで共通の意味をもつスコアを算出できる。ただし等化（後述）が必要。

項目母数によるテスト版のカスタマイズ

《古典的テスト理論》

- 「I-T相関係数」「正答率」「設問回答率分析図」などを用いて，事後的に項目の評価を行う
- 得られる評価指標は受験者集団に依存しているため，他の受験者集団に対しても同様の意味をもつ指標になるとは限らない

《項目反応理論》

- 推定された項目母数は受験者集団に依存しない
- あらかじめ項目母数を推定しておけば，それらの項目を利用して，特定の潜在特性値にターゲットを絞ったテスト版を作成することが可能
- 項目母数からテスト情報関数（後述）を求めることで，テスト版がどの程度の潜在特性値をもつ受験者向けであるかを知ることができる

項目反応理論による測定精度の指標

《古典的テスト理論》

- 信頼性係数（アルファ係数など）
- 測定の標準誤差（Standard Error of Measurement, SEM）

- テスト版一つに対し，一つの値が推定される。
- 受験者の能力分布に依存している。

《項目反応理論》

- 項目情報関数
- テスト情報関数

- 項目情報関数は，項目単位で推定される。
- 受験者の能力分布に依存していない。
- 受験者の能力別に情報量（推定精度）の値が求まる。

項目反応理論を用いる理由（まとめ）

《古典的テスト理論》

- 項目特性（困難度）と受験者の能力が交絡しているため，これらを区別して考えることができない
- 項目困難度や，テスト版を受験した集団に依存した指標を用いて受験者の成績を表す（偏差値を用いた場合も同様）

《項目反応理論》

- 項目特性（困難度など）と受験者の能力を分けてモデル上で表現している
- テスト版の項目特性に依存しない「潜在特性値」を用いて受験者の成績を表す

- 項目反応理論(IRT)を用いる理由
- IRTの概要
  - IRTによる尺度化
  - 母数の推定方法
  - 母数の推定精度とテスト情報関数
  - 実践例
  - 項目反応モデルの拡張
- IRTによる等化の概要
  - 等化の必要性
  - 等化の方法

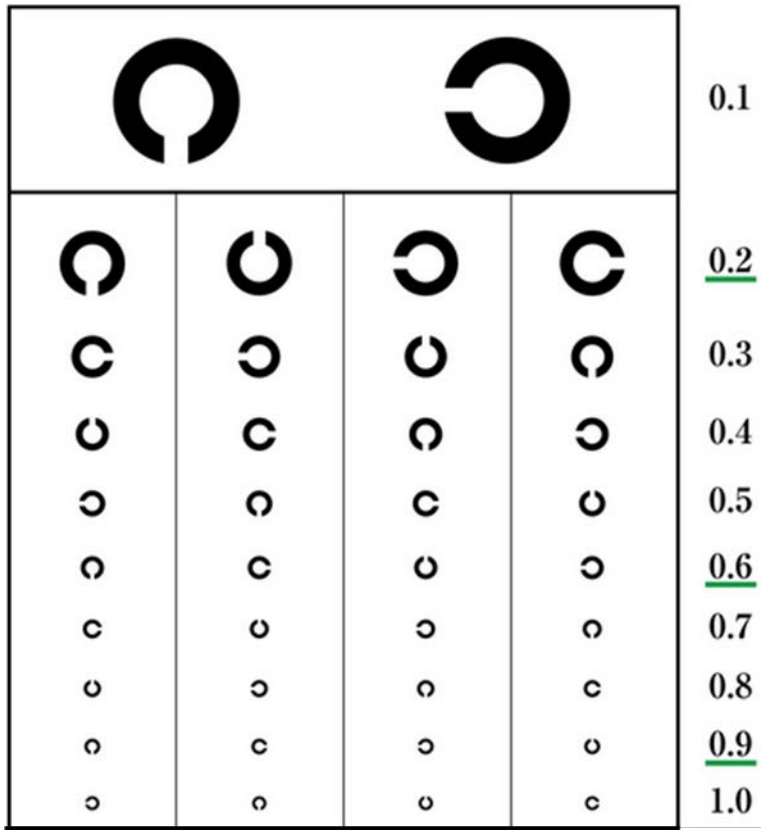
正誤データ(0-1データ) u_{ij} $\rightarrow j$

i:受験者
j:項目

	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9	Q10	正答数(素点)
A	1	0	1	1	0	1	1	1	1	1	8
B	1	1	1	1	0	1	1	1	1	1	9
C	1	1	0	0	1	1	1	1	1	0	7
D	0	1	0	1	0	1	1	1	1	0	6
E	0	0	0	0	1	0	0	0	0	0	1
F	1	0	1	1	1	1	0	0	0	0	5
G	1	1	1	0	1	0	0	0	0	0	4
H	1	1	1	1	0	0	0	0	0	0	4
I	1	1	1	1	1	1	1	1	0	1	9
J	1	1	1	0	0	0	0	0	0	0	3
K	1	1	0	0	1	1	1	1	1	1	8
L	1	0	0	0	1	0	0	0	0	0	2
正答数	10	8	7	6	7	7	7	7	5	4	
正答率(通過率)	.83	.67	.58	.50	.58	.58	.58	.58	.42	.33	

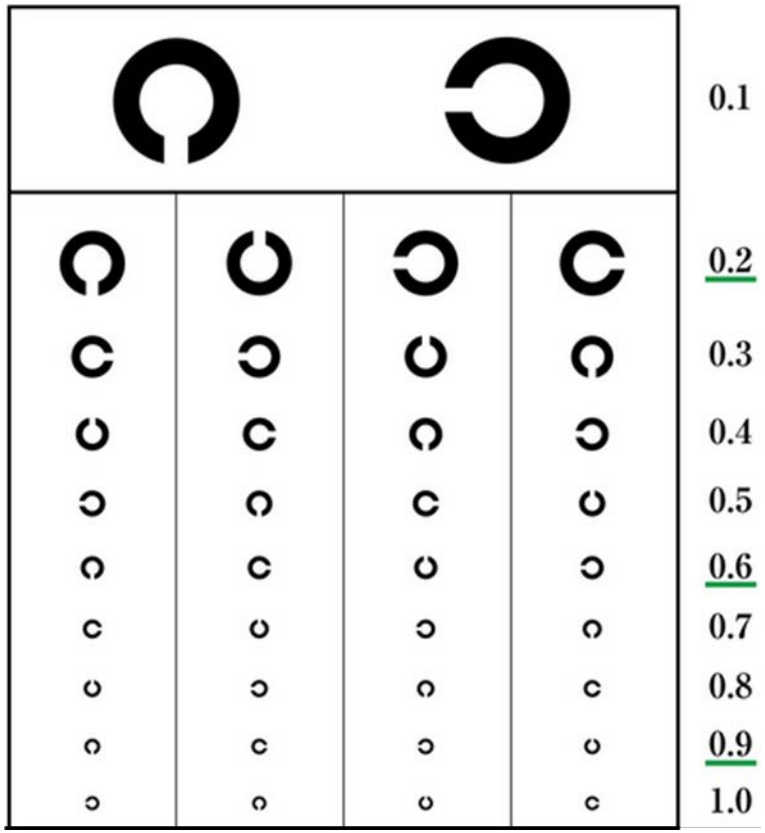
↓ i

視力検査のデータ



	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0
A	1	1	1	1	1	1	1	0	0	0
B	1	1	1	1	1	1	1	1	1	1
C	1	1	1	1	1	1	1	1	1	1
D	1	1	1	1	1	1	0	0	0	0
E	1	0	0	0	0	0	0	0	0	0
F	1	1	1	1	1	1	1	1	1	0
G	1	1	1	0	0	0	0	0	0	0
H	1	1	1	1	0	0	0	0	0	0
I	1	1	1	1	1	0	0	0	0	0
J	1	1	0	0	0	0	0	0	0	0
K	1	1	1	1	0	0	0	0	0	0
L	1	1	1	1	1	1	1	1	1	0

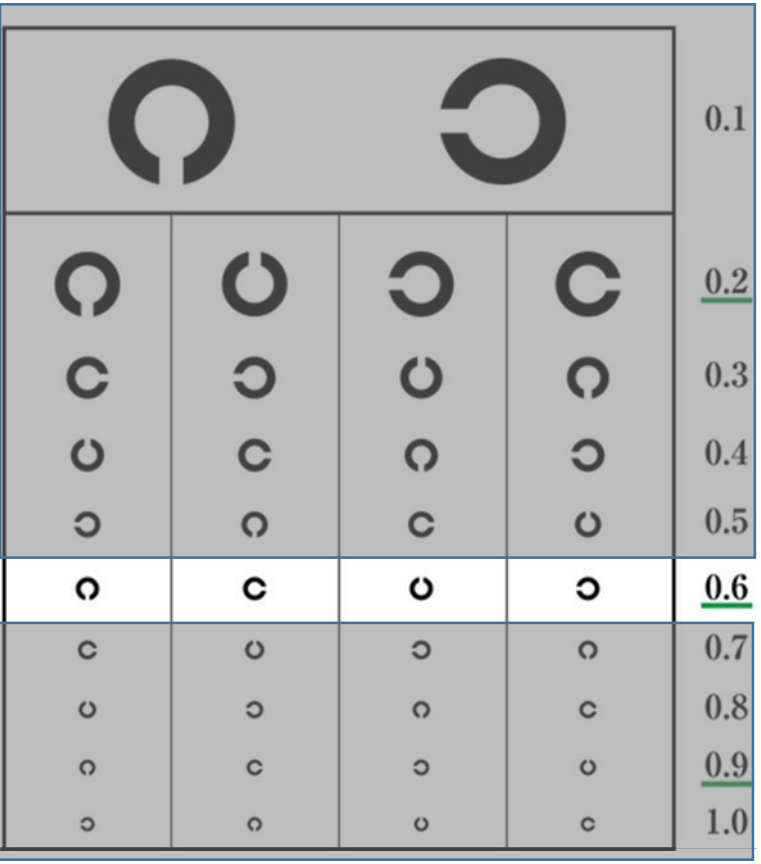
ガットマンスケール



	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0
B	1	1	1	1	1	1	1	1	1	1
C	1	1	1	1	1	1	1	1	1	1
F	1	1	1	1	1	1	1	1	1	0
L	1	1	1	1	1	1	1	1	1	0
A	1	1	1	1	1	1	1	0	0	0
D	1	1	1	1	1	1	0	0	0	0
I	1	1	1	1	1	0	0	0	0	0
H	1	1	1	1	0	0	0	0	0	0
K	1	1	1	1	0	0	0	0	0	0
G	1	1	1	0	0	0	0	0	0	0
J	1	1	0	0	0	0	0	0	0	0
E	1	0	0	0	0	0	0	0	0	0

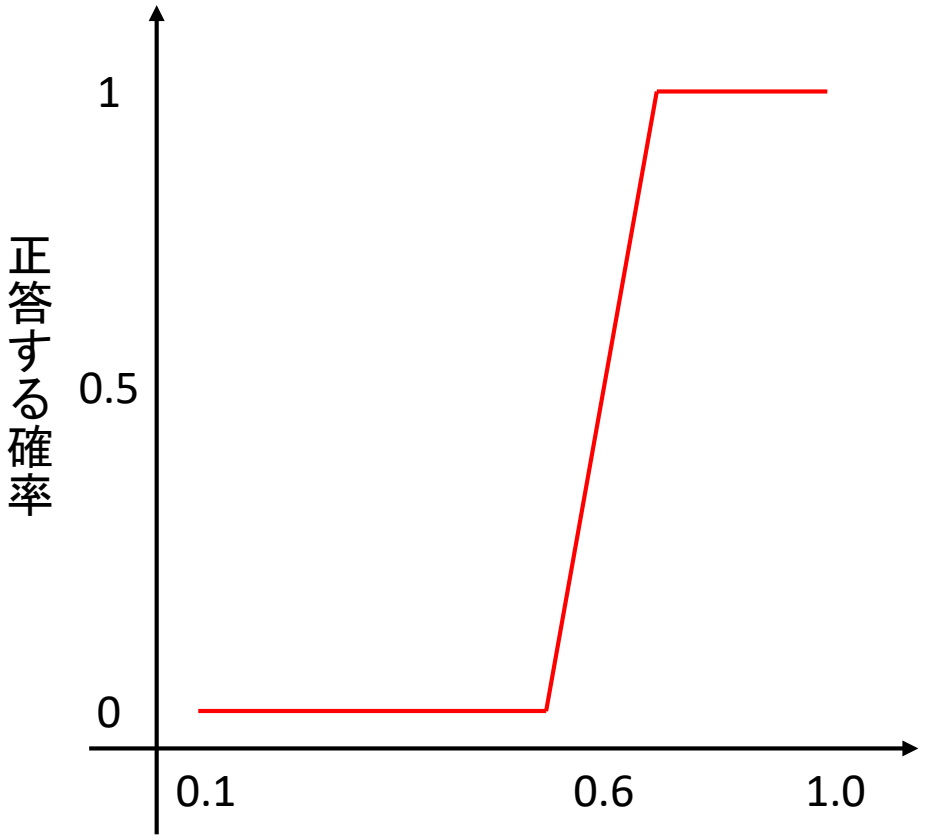
左上に「1」が並ぶように並べ替え

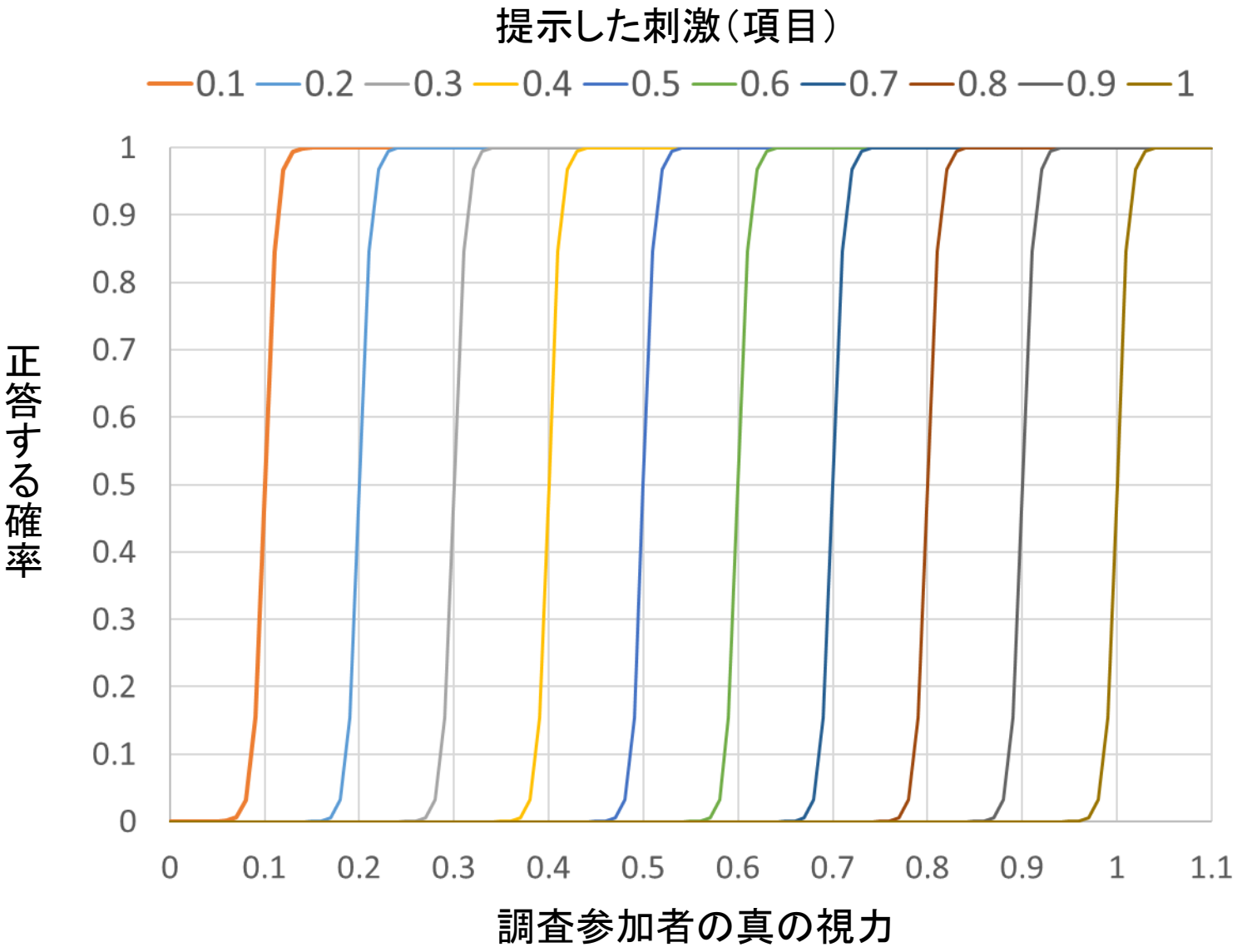
さまざまな視力をもつ者からデータをとる



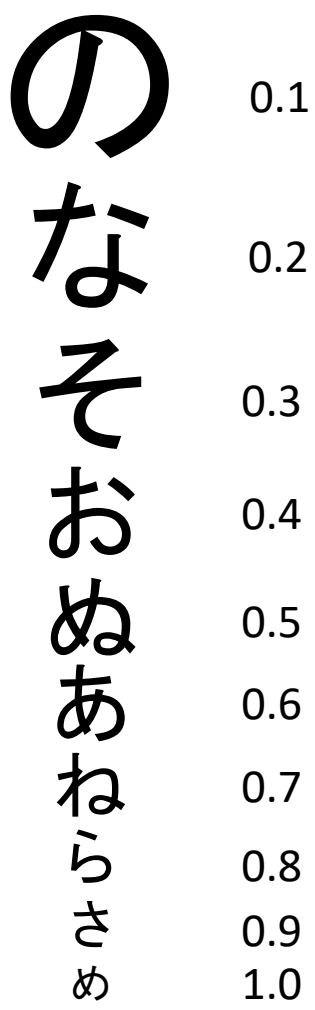
調査参加者の視力

	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0
正	0	0	0	0	0	5	10	10	10	10
誤	10	10	10	10	10	5	0	0	0	0



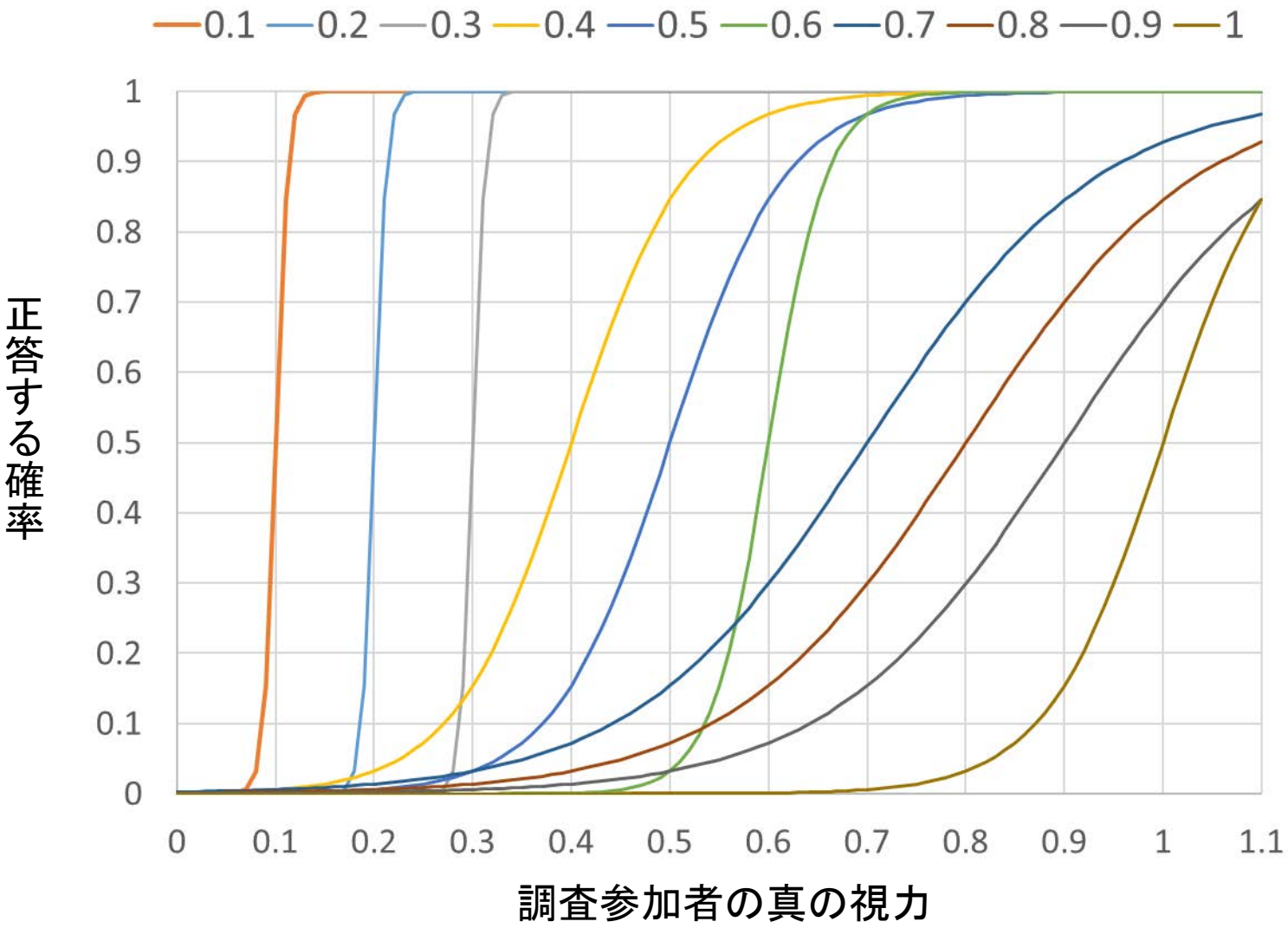

提示した刺激（項目）
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
正答する確率
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
1.1
調査参加者の真の視力

ガットマンスケール（識別力の低い場合）



	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0
B	1	1	1	1	1	0	1	1	0	1
C	1	1	1	1	1	1	0	1	0	0
F	1	1	1	1	1	1	0	0	1	0
L	1	1	1	1	1	0	1	1	1	0
A	1	1	1	1	1	0	1	0	0	0
D	1	1	1	1	0	1	0	1	0	0
I	1	1	1	0	1	0	1	1	0	0
H	1	1	1	1	0	1	0	0	0	0
K	1	1	1	1	0	0	1	0	1	0
G	1	1	1	0	0	0	0	0	0	1
J	1	1	0	0	0	0	0	0	0	0
E	1	0	0	0	0	0	0	0	0	0

文字の形が互いに似通っている場合，識別が難しくなる


0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
正答する確率
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
1.1
調査参加者の真の視力

ガットマンスケール

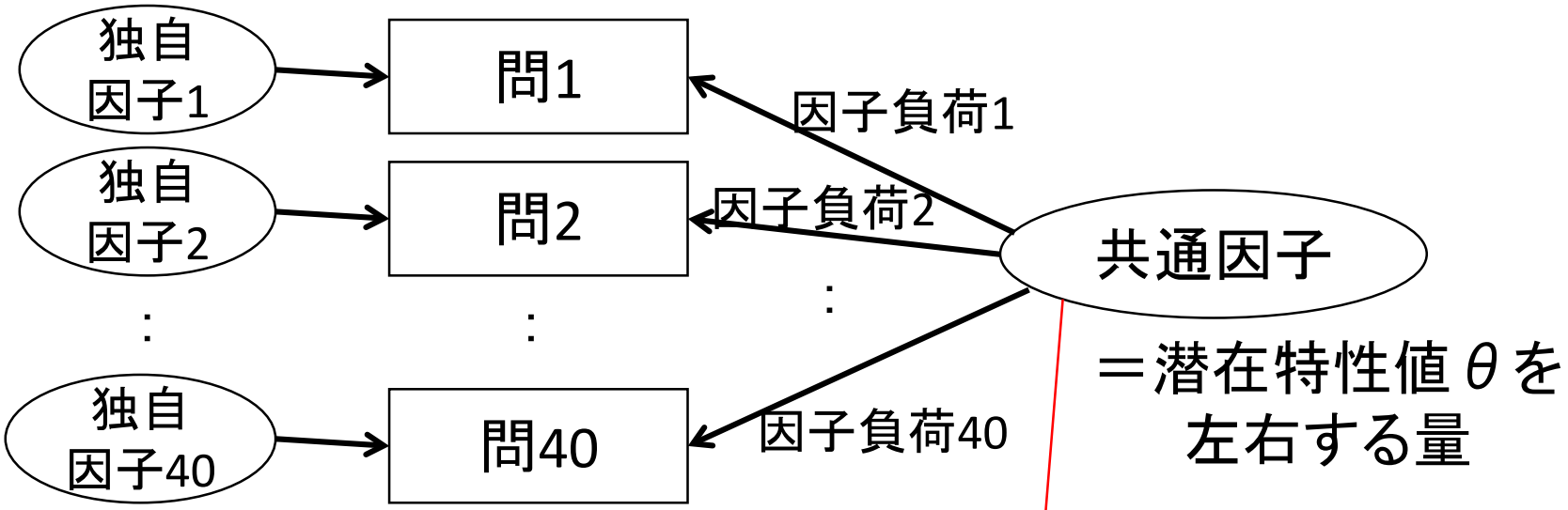
- ガットマンスケールによって、「物理的な大小を心理量に変換する」ための「尺度」を作成することができる。
  - 刺激（＝ランドルト環）の大きさ（物理量）→視力（心理量）
- 刺激に対して正答できるかどうかの「確率」と，調査参加者の「真の視力」との対応関係を記述することで，それぞれの刺激が調査参加者の視力をどの程度識別できるかの指標を与えることができる。

- 調査参加者の「真の視力」は観測できない。
- 学力テストの場合，測定される構成概念は複数考えられる。
  - 「テストで測定される構成概念は一つ」と仮定

項目反応理論（項目応答理論，item response theory; IRT）

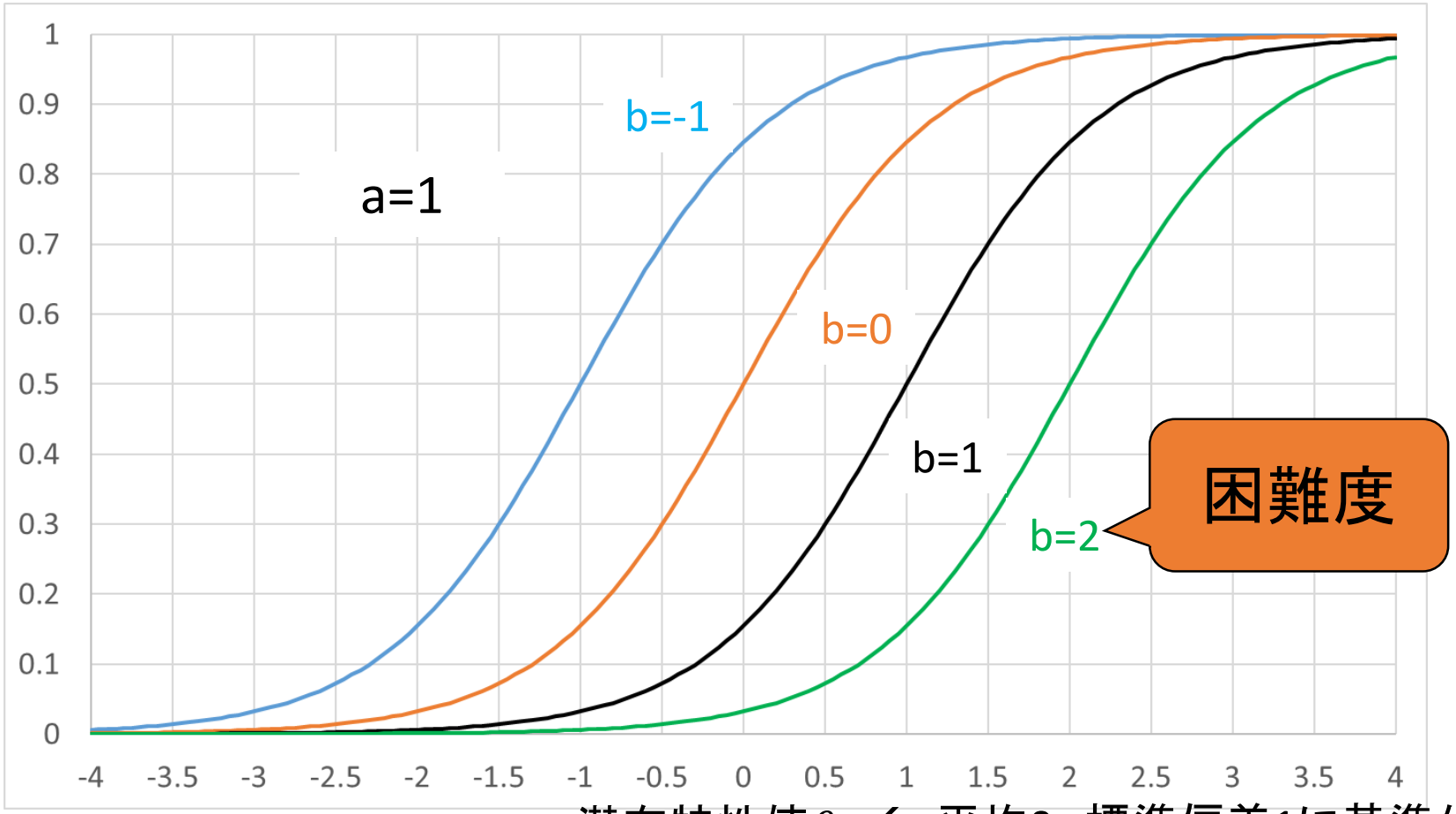
- 受験者がある項目に正答する確率を記述する。
- 「潜在特性値」（能力値）によって，正答確率を説明するモデル。

- 「潜在特性値θ」・・・「正答・非正答」のデータから抽出される
  1因子の因子得点



正誤反応の項目間相関から抽出できる

項目特性曲線（ICC）（1パラメタ・ロジスティックモデル）



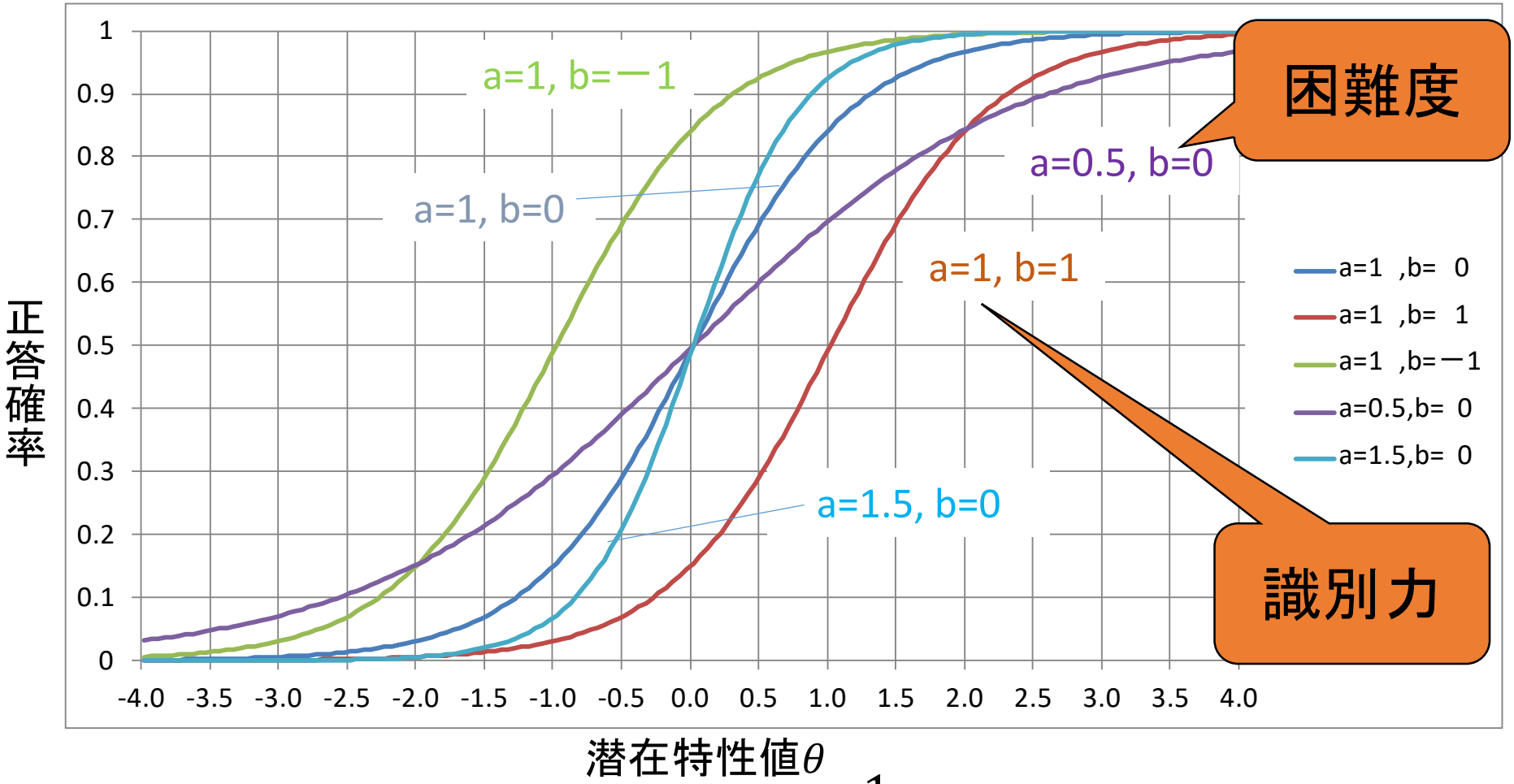
潜在特性値θ　← 平均0，標準偏差1に基準化

受験者集団を規準集団として，標準化得点（Z得点）と同様の解釈が可能に

正答確率と潜在特性値を，IRTのモデルに基づき記述した式

$$P_j(\theta|a,b_j)=\frac{1}{1+\exp(-Da(\theta-b_j))}$$

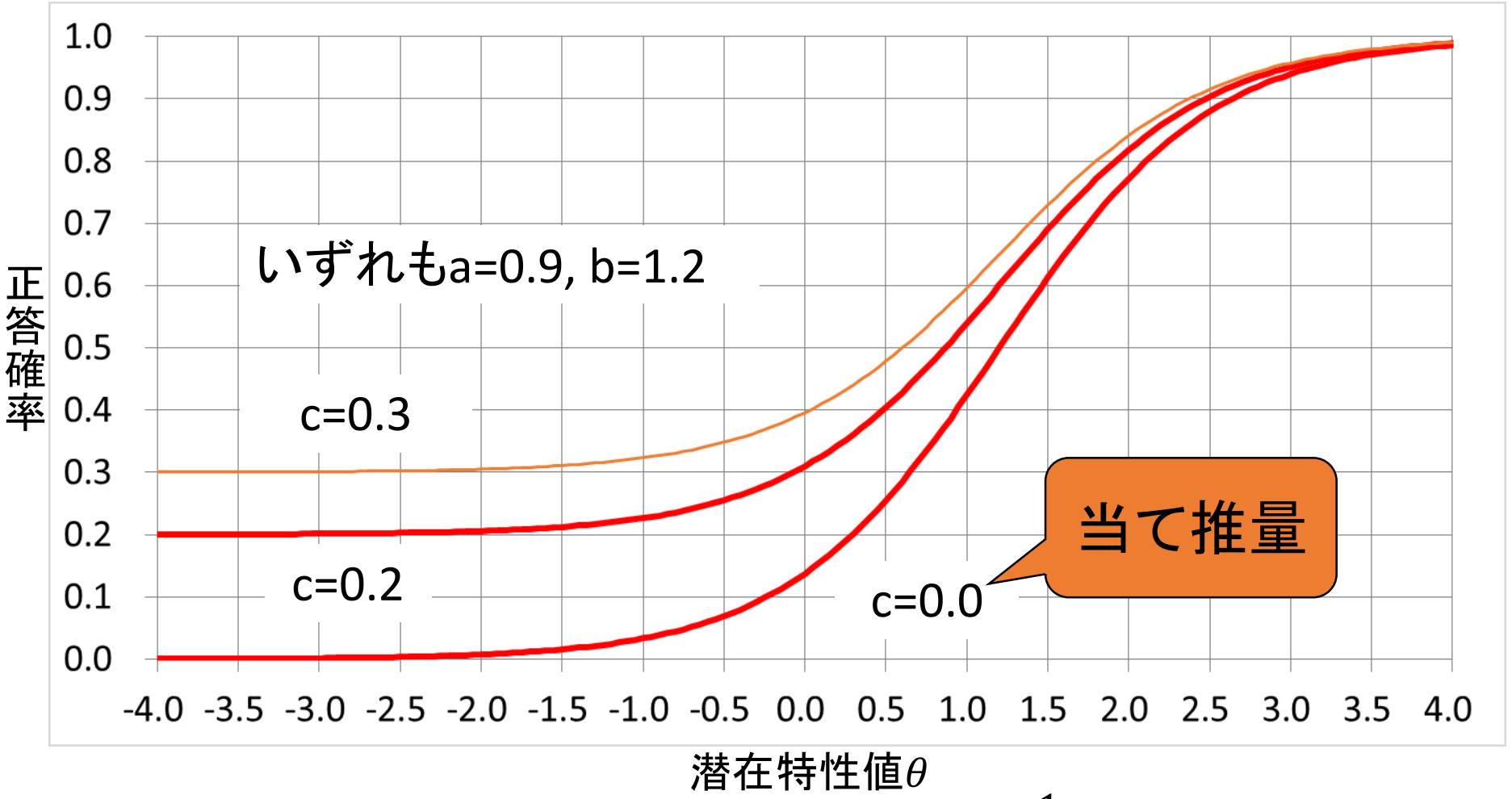
ICC（2パラメタ・ロジスティックモデル）



$$P_j(\theta|a_j, b_j) = \frac{1}{1 + \exp(-Da_j(\theta - b_j))}$$

困難度 = 正答確率50%となる潜在特性値

ICC（3パラメタ・ロジスティックモデル）



$$P_j(\theta|a_j, b_j, c_j) = c_j + (1 - c_j)\frac{1}{1 + \exp(-Da_j(\theta - b_j))}$$

c が0以外の場合，困難度 b = 正答確率50%となる潜在特性値とはならない

IRTで仮定されるモデル（2値型）

1パラメタ・ロジスティックモデル

$$P_j(\theta|a, b_j) = \frac{1}{1 + \exp(-Da(\theta - b_j))}$$

2パラメタ・ロジスティックモデル

$$P_j(\theta|a_j, b_j) = \frac{1}{1 + \exp(-Da_j(\theta - b_j))}$$

3パラメタ・ロジスティックモデル

$$P_j(\theta|a_j, b_j, c_j) = c_j + (1 - c_j)\frac{1}{1 + \exp(-Da_j(\theta - b_j))}$$

a_j: 識別力母数　　b_j:困難度母数　　c_j: 当て推量母数
（下方漸近パラメタ）

D: 尺度因子（$= 1.7$）

$P_j(\theta|a_j, b_j)$:　識別力母数・困難度母数で条件づけられた，ある潜在特性値における正答確率

$\exp(x)$:　e^xを表し，eは自然対数の底（=2.718281828……）

局所独立の仮定／テストの一次元性

潜在特性値θをある値に固定した場合に，テストに含まれる項目に対する解答は相互に独立に生ずる，という仮定。

u_j ：正誤データ：正解なら「1」，不正解なら「0」の値をとる．

n項目からなるテストに対する解答が$\boldsymbol{u} = \{u_1, u_2 \ldots u_j \ldots u_n\}$であった場合，解答データが得られる確率は以下の式で表される。

$$P\left(u_1, u_2 \ldots u_j \ldots u_n \middle| \theta\right) = \prod_{j=1}^{n} P_j(\theta)^{u_j}\left(1 - P_j(\theta)\right)^{1-u_j}$$

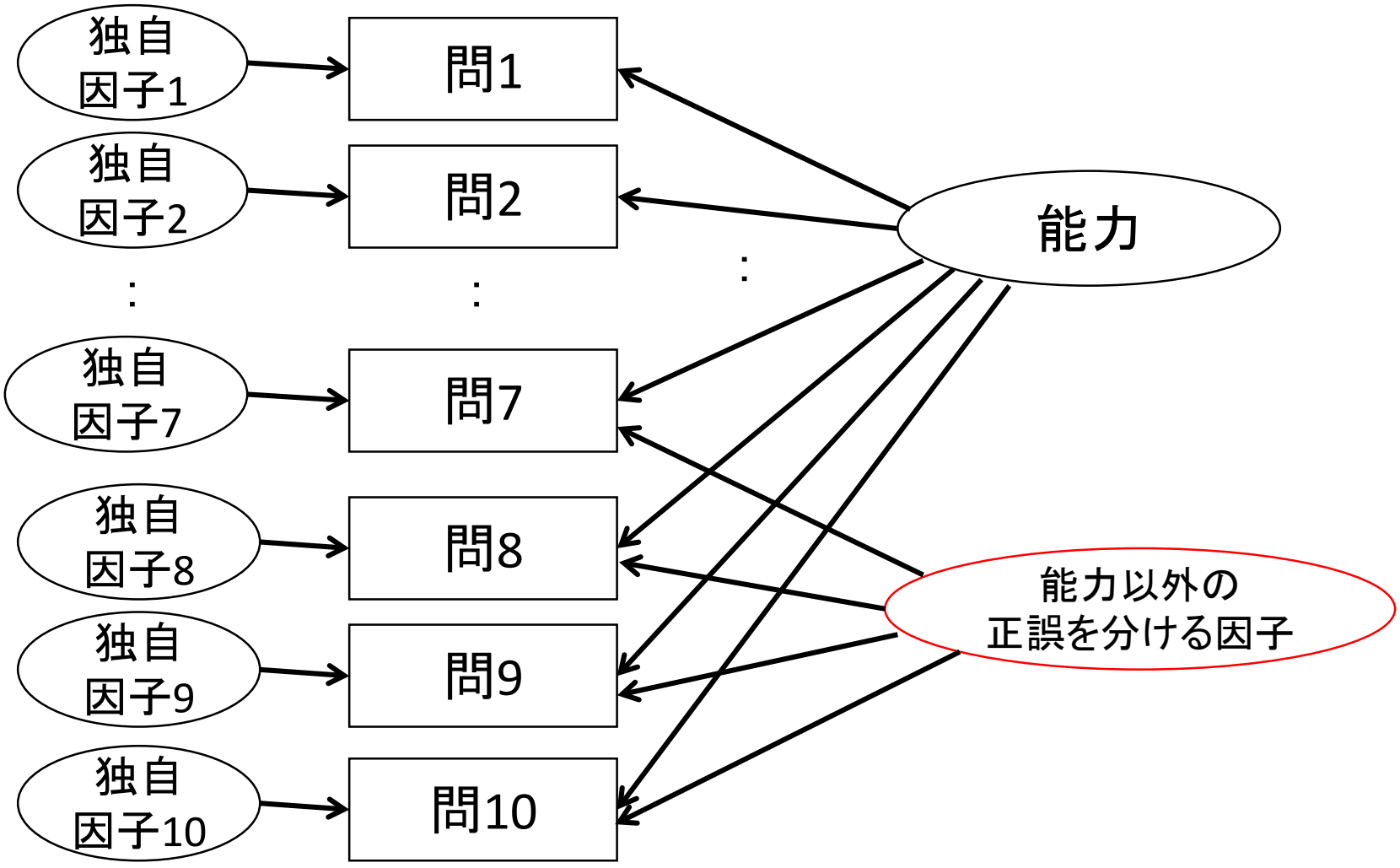
例：$P\left(u_1, u_2 \middle| \theta\right) = \left(P_1(\theta)^{u_1}(1 - P_1(\theta))^{1-u_1}\right) \times \left(P_2(\theta)^{u_2}(1 - P_2(\theta))^{1-u_2}\right)$

この例で$\{u_1, u_2\} = \{1,0\}$の場合，　$P(u_1, u_2|\theta) = P_1(\theta) \times (1 - P_2(\theta))$

「データが得られる確率」（θの条件付き）が，ICCの積の形になっていることに注意。

テストに含まれる項目群に共通した潜在特性がひとつであるということを意味している（＝テストの一次元性を表す）

局所独立の仮定／テストの一次元性が満たされない場合



局所独立の仮定／テストの一次元性が満たされない場合の具体例

【問4-3】
(1) 1188の正の約数は
（　）個ある。
(2) そのうち，2の倍数は
（　）個，4の倍数は
（　）個ある。
(3) 1188のすべての正の約数の積を2進法で表すと，末尾には0が
（　）個並ぶ。

（2017年度大学入試センター試験「数学I・数学A」p.39を筆者改変）

- ある項目への正答が，他の項目に正答するための条件となっている場合
  - 問1に正答できることが，問2に正答するために必要
  - 大きな「問」の下に，小さな設問が多数存在する場合に起こりやすい（大問形式）

  対処法：
  - 項目の構成を，独立して回答できる形の「小問独立形式」とする
  - 大問全体を「テストレット」とみなして，正答数をカウントし，多値型モデル（後述）を適用する

- 複数の構成概念によって正誤が決まるような項目の場合

  対処法：
  - 一つの問題で測定される構成概念は一つに絞るように設問を工夫する

- 項目反応理論(IRT)を用いる理由
- **IRTの概要**
  - IRTによる尺度化
  - **母数の推定方法**
  - 母数の推定精度とテスト情報関数
  - 実践例
  - 項目反応モデルの拡張
- IRTによる等化の概要
  - 等化の必要性
  - 等化の方法

母数推定(項目母数)

- 正誤データを用いて，項目母数を推定。
- ある程度多くの受験者を必要とする。
- 項目反応モデルは，分析者が指定する。
- モデルの情報量規準を算出することで，最適な項目反応モデルを選択する手がかりとなる。

推定方法

- 同時最大尤度法：項目母数と潜在特性値を同時に推定する方法。あまり用いられない
- 周辺最大尤度法：潜在特性値を周辺分布とおき，項目母数を推定する方法。よく用いられる
  - EMアルゴリズムによる最適化
- 他に，ベイズ統計学の枠組みで母数を推定する方法も近年用いられつつある。
- 方法についての詳細は加藤・山田・川端(2014)などを参照

母数推定の流れ（周辺最尤法による項目母数の推定→潜在特性値の推定）

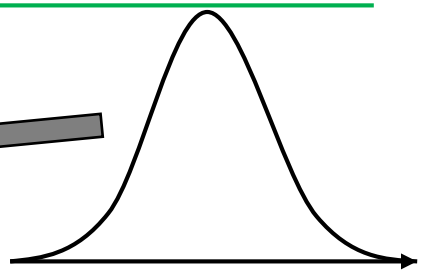
①
項目母数の
推定

正誤データ

	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9	Q10
A	1	0	1	1	0	1	1	1	1	1
B	1	1	1	1	0	1	1	1	1	1
C	1	1	0	0	1	1	1	1	1	0
D	0	1	0	1	0	1	1	1	1	0
E	0	0	0	0	1	0	0	0	0	0
F	1	0	1	1	1	1	0	0	0	0
G	1	1	1	0	1	0	0	0	0	0
H	1	1	1	1	0	0	0	0	0	0
I	1	1	1	1	1	1	1	1	0	1
J	1	1	1	0	0	0	0	0	0	0
K	1	1	0	0	1	1	1	1	1	1
L	1	0	0	0	1	0	0	0	0	0

項目母数の推定値

	識別力	困難度
Q1	0.8	0.0
Q2	1.0	0.8
:	:	:
Q10	1.1	0.6



潜在特性値の分布を用いて推定するが，潜在特性値そのものは推定しない（局外母数）

②
潜在特性値
の
推定

正誤データ

	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9	Q10
A	1	0	1	1	0	1	1	1	1	1
B	1	1	1	1	0	1	1	1	1	1
C	1	1	0	0	1	1	1	1	1	0
D	0	1	0	1	0	1	1	1	1	0
E	0	0	0	0	1	0	0	0	0	0
F	1	0	1	1	1	1	0	0	0	0
G	1	1	1	0	1	0	0	0	0	0
H	1	1	1	1	0	0	0	0	0	0
I	1	1	1	1	1	1	1	1	0	1
J	1	1	1	0	0	0	0	0	0	0
K	1	1	0	0	1	1	1	1	1	1
L	1	0	0	0	1	0	0	0	0	0

項目母数の推定値

	識別力	困難度
Q1	0.8	0.0
Q2	1.0	0.8
:	:	:
Q10	1.1	0.6

	潜在特性値
A	0.8
B	1.0
C	-1.2
:	:

母数推定の詳細(潜在特性値;最尤推定)

項目母数がわかっている場合に、潜在特性値を推定するには？

	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9	Q10
a_j	0.5	0.8	0.7	0.5	1.2	1.2	1.1	0.8	0.5	0.4
b_j	-1.5	-1.2	2.0	2.2	0.8	0.6	-0.2	1.8	-1.1	-0.4

生徒	問1	2	3	4	5	6	7	8	9	10
1	1	0	1	1	1	1	1	1	1	1
2	1	1	1	1	1	1	1	1	0	1
3	1	0	1	1	1	1	0	1	1	0
4	1	1	1	1	1	1	1	0	1	0
5	1	1	0	1	1	0	1	1	1	0
6	1	0	0	1	1	1	1	1	1	0
7	1	1	1	1	1	0	1	1	0	0
8	1	1	1	0	0	1	1	0	0	1
9	1	0	1	0	1	1	0	0	1	1
10	1	0	0	0	0	1	1	1	1	0
11	1	1	1	0	0	0	0	1	0	1
12	1	1	0	1	0	1	0	0	0	1
13	1	0	0	1	1	0	0	1	0	0
14	1	1	1	0	0	1	0	0	0	1
：		：						：		
1000	0	1	1	1	0	0	1	1	1	0

← $\boldsymbol{u}_1 = \{1,0,1,1,1,1,1,1,1,1\}$

推定は受験者ごとに
独立して行う。

必要な情報：
- 項目母数
- 正誤反応ひとり分 u_j

母数推定（潜在特性値；最尤推定）

ICC（2PL）

$$P_j(\theta) = \frac{1}{1 + \exp(-Da_j(\theta - b_j))}$$

項目母数のセット

尤度関数

$$L = \prod_{j=1}^{n} P_j(\theta)^{u_j}\left(1 - P_j(\theta)\right)^{1-u_j}$$

正誤反応

対数尤度関数

$$\ln L = \sum_{j=1}^{n}\left[u_j \ln P_j(\theta) + (1 - u_j)\ln(1 - P_j(\theta))\right]$$

尤度方程式

$$\frac{\partial \ln L}{\partial \theta} = 0$$

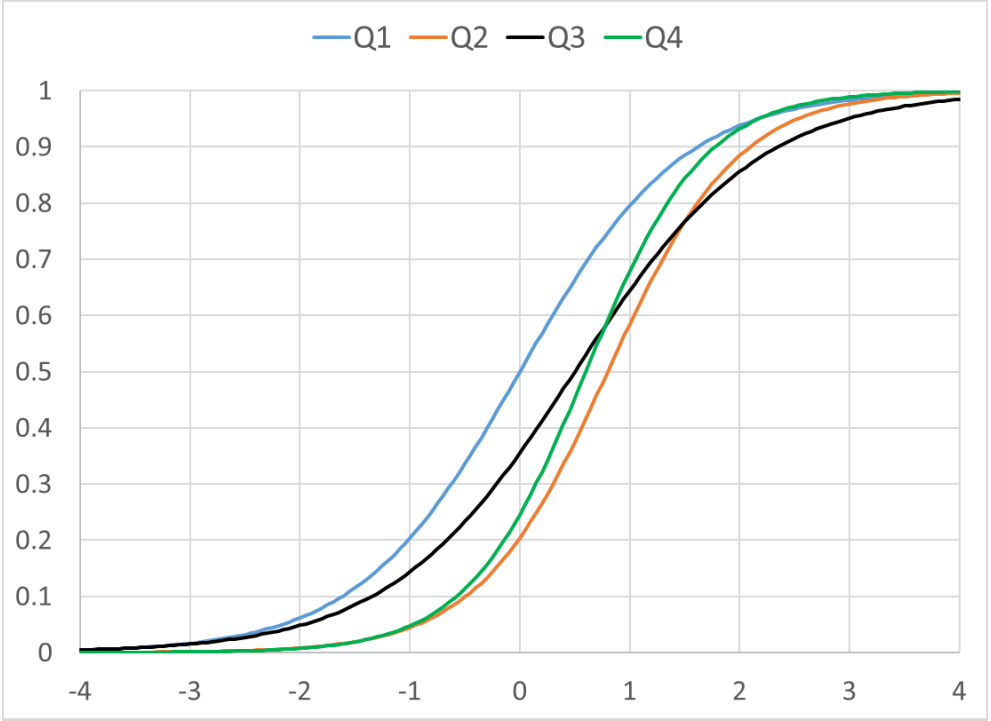
通常はNewton-Raphson法などにより数値的に求める.

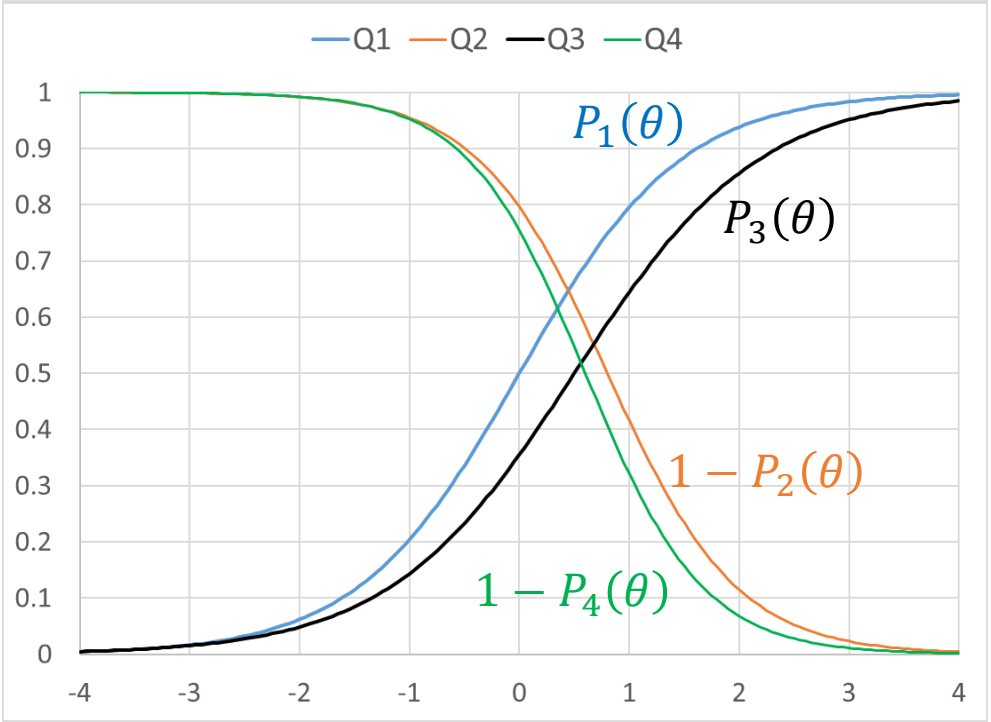
	識別力	困難度	正誤u_j
1	0.8	0.0	1
2	1.0	0.8	0
3	0.7	0.5	1
4	1.1	0.6	0

$$L = P(u_1, u_2 \dots u_j \dots u_n | \theta)$$

$$= \prod_{j=1}^{n} P_j(\theta)^{u_j} \left(1 - P_j(\theta)\right)^{1-u_j}$$

$$= P_1(\theta) \times \left(1 - P_2(\theta)\right) \times P_3(\theta) \times \left(1 - P_4(\theta)\right)$$

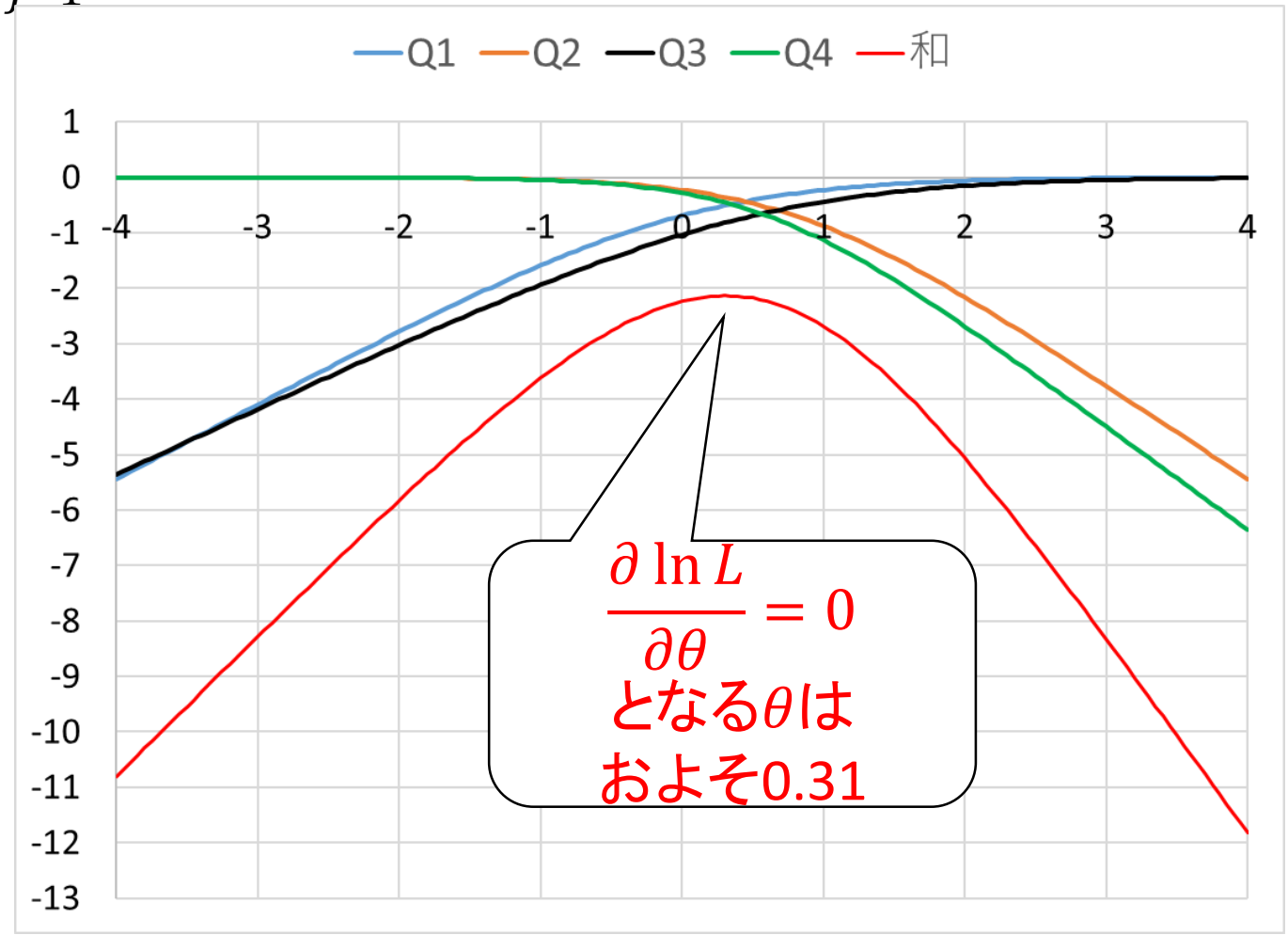




$$\ln L = \sum_{j=1}^{n} \left[u_j \ln P_j(\theta) + (1 - u_j) \ln(1 - P_j(\theta)) \right]$$

極値付近の接線の傾きは0となる

$\ln L$を微分して0となるθを求める。



$\ln L$の最大値を返すθが，最尤推定値。

最大化された$\ln L$は，AIC（Akaike Information Criterion）など，モデルの情報量規準を計算するために用いられる。

- 項目反応理論(IRT)を用いる理由
- **IRTの概要**
  - IRTによる尺度化
  - 母数の推定方法
  - **母数の推定精度とテスト情報関数**
  - 実践例
  - 項目反応モデルの拡張
- IRTによる等化の概要
  - 等化の必要性
  - 等化の方法

θの推定の精度

	識別力	困難度
1	0.8	0.0
2	1.0	0.8
3	0.7	0.5
4	1.1	0.6

これら4項目に対して、潜在特性値θが0.6 の受験者が1000名受験したところ、潜在特性値の推定結果$\hat{\theta}$の平均は0.59，標準偏差は0.43であった。

それに対し、θが1.2 の受験者が1000名受験したところ、$\hat{\theta}$の平均は1.09，標準偏差は0.88であった。

推定誤差

$$e_i = \hat{\theta}_i - \theta_i$$

$\hat{\theta}$の分散が大きい＝推定誤差が大きいことを示している。

精度良く推定できるθとは，具体的にどのθの範囲であろうか？

古典的テスト理論での「測定の標準誤差」は，一つのテスト版に対して一つだけ推定される。

最尤推定の性質

- 最尤推定値の標本分布は、項目数が大きくなるにしたがって正規分布に近づく（漸近正規性）

- 最尤推定値の平均は、項目数を増大させた場合、真の値に近づく（漸近不偏性）。

$$E[\hat{\theta}_i|\theta_i] \to \theta_i$$

- 最尤推定値の分散は、項目数を増大させた場合、以下の値に近づく（漸近有効性）。

$$V[\hat{\theta}_i|\theta_i] \to \frac{1}{I(\theta_i)}$$

$$I(\theta_i) = -E\left[\left(\frac{\partial}{\partial\theta}\ln L(\theta|u_{ij})\right)^2_{\theta=\theta_i}\right]$$

（フィッシャー情報量）

推定値の誤差分散から，推定の精度が評価できる

テスト情報関数$I(\theta_i)$と推定値の標準誤差

2PLの場合

（1PLの場合はすべてのjについて$a_j = 1$とおく）

$$I(\theta_i) = \sum_{j=1}^{n} D^2 a_j^2 P_j(\theta_i)\left(1 - P_j(\theta_i)\right)$$

項目ごとの情報量$I_j(\theta_i)$
＝項目情報関数

3PLの場合

$$I(\theta_i) = \sum_{j=1}^{n} \frac{D^2 a_j^2 \left(P_j(\theta_i) - c_j\right)^2 \left(1 - P_j(\theta_i)\right)}{P_j(\theta_i)(1 - c_j)^2}$$

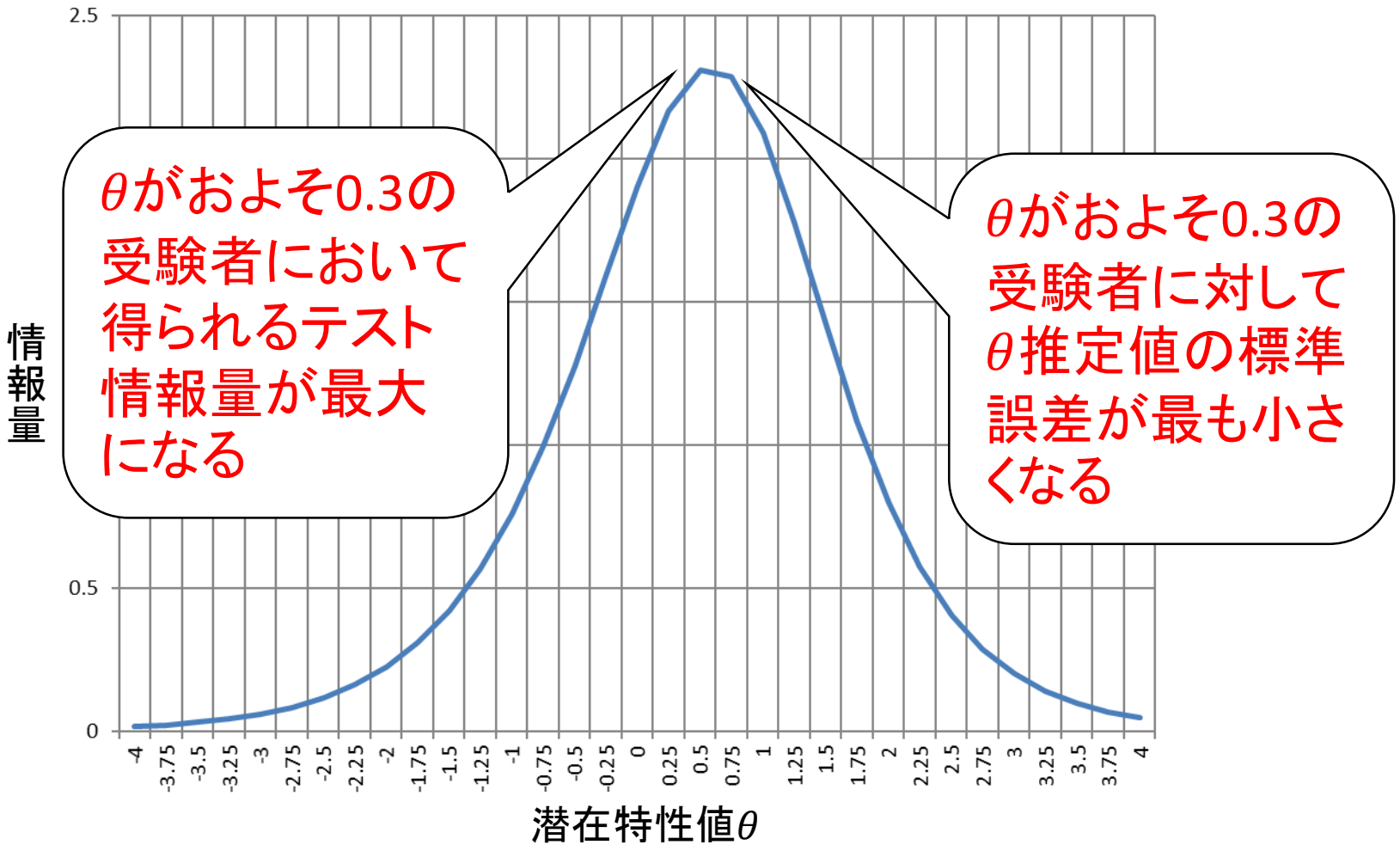
ある潜在特性値θ_iをもつ受験者iにおける推定値$\widehat{\theta_i}$の標準誤差

$$SE = \frac{1}{\sqrt{I(\theta_i)}}$$

θの95%信頼区間

$$\widehat{\theta_i} - 1.96 \times SE \leq \theta \leq \widehat{\theta_i} + 1.96 \times SE$$

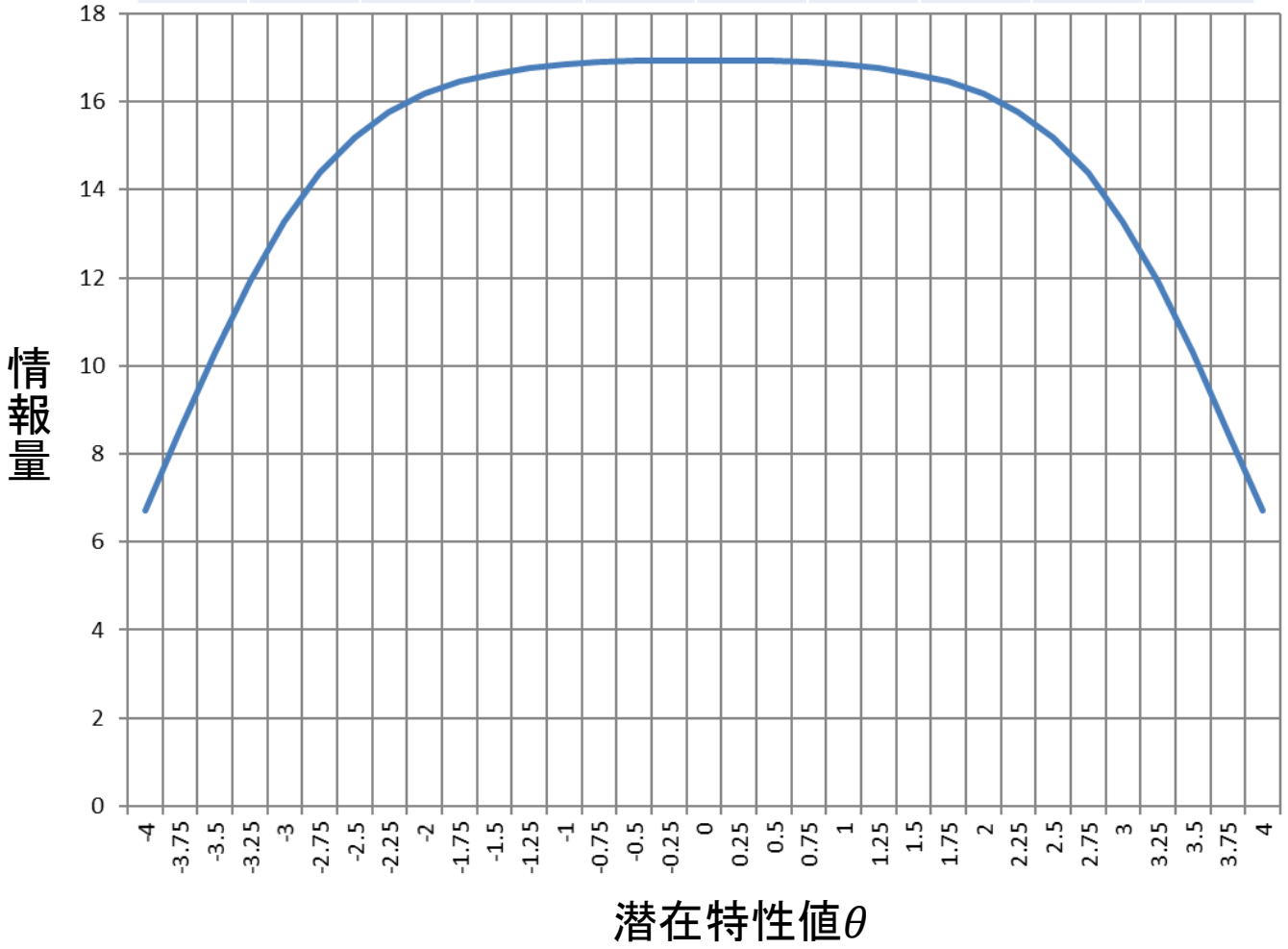
θと情報量との関係（テスト情報関数）



項目ごとに$I(\theta)$を算出してその和をとり，θの値ごとにプロットしたもの

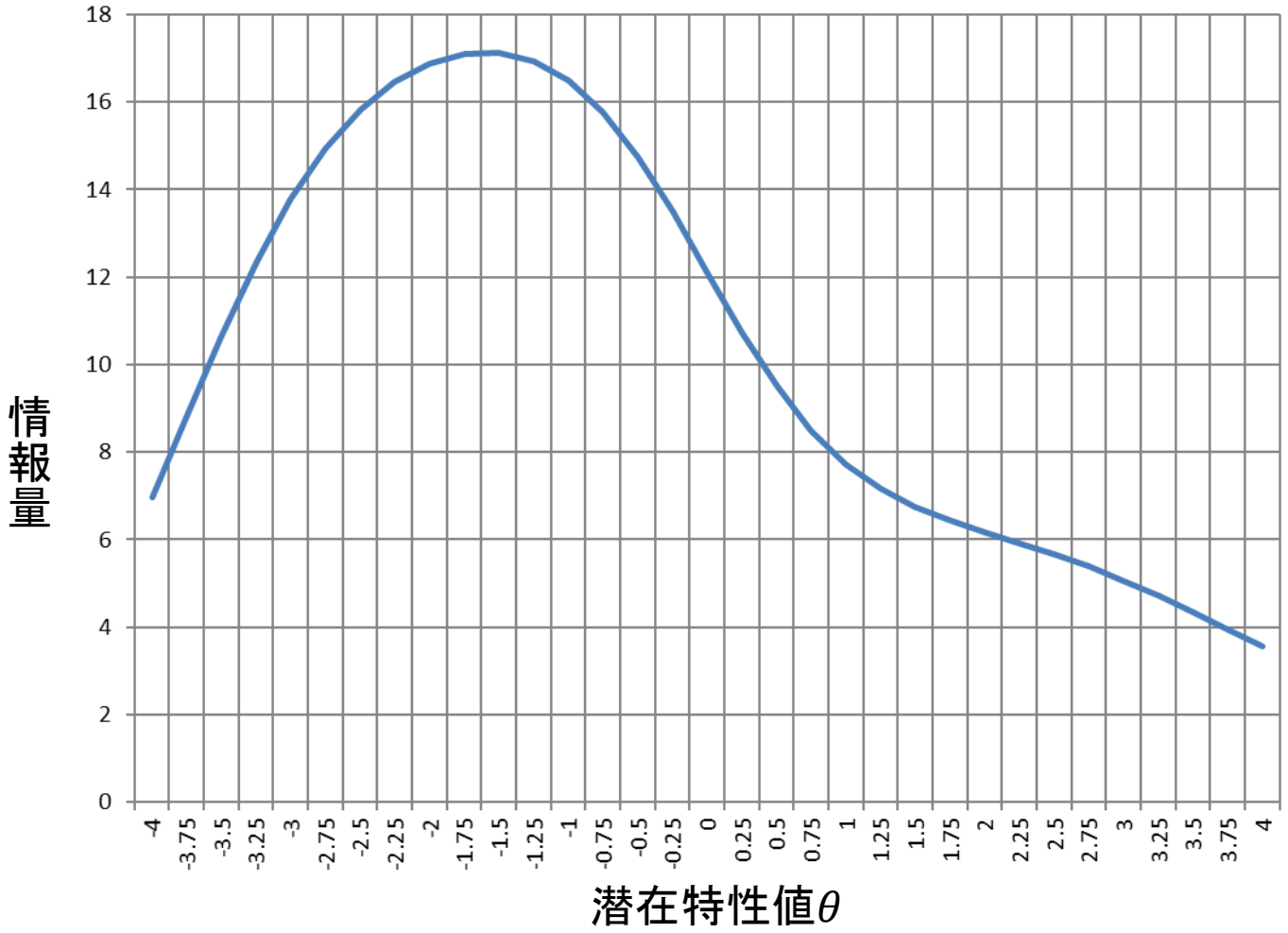
すべての項目で識別力が同じテスト版（1PL）

問	1	2	……	37	38	39	……	74	75
a_j	1.0	1.0	……	1.0	1.0	1.0	……	1.0	1.0
b_j	-3.7	-3.6	……	-0.1	0.0	0.1	……	3.6	3.7

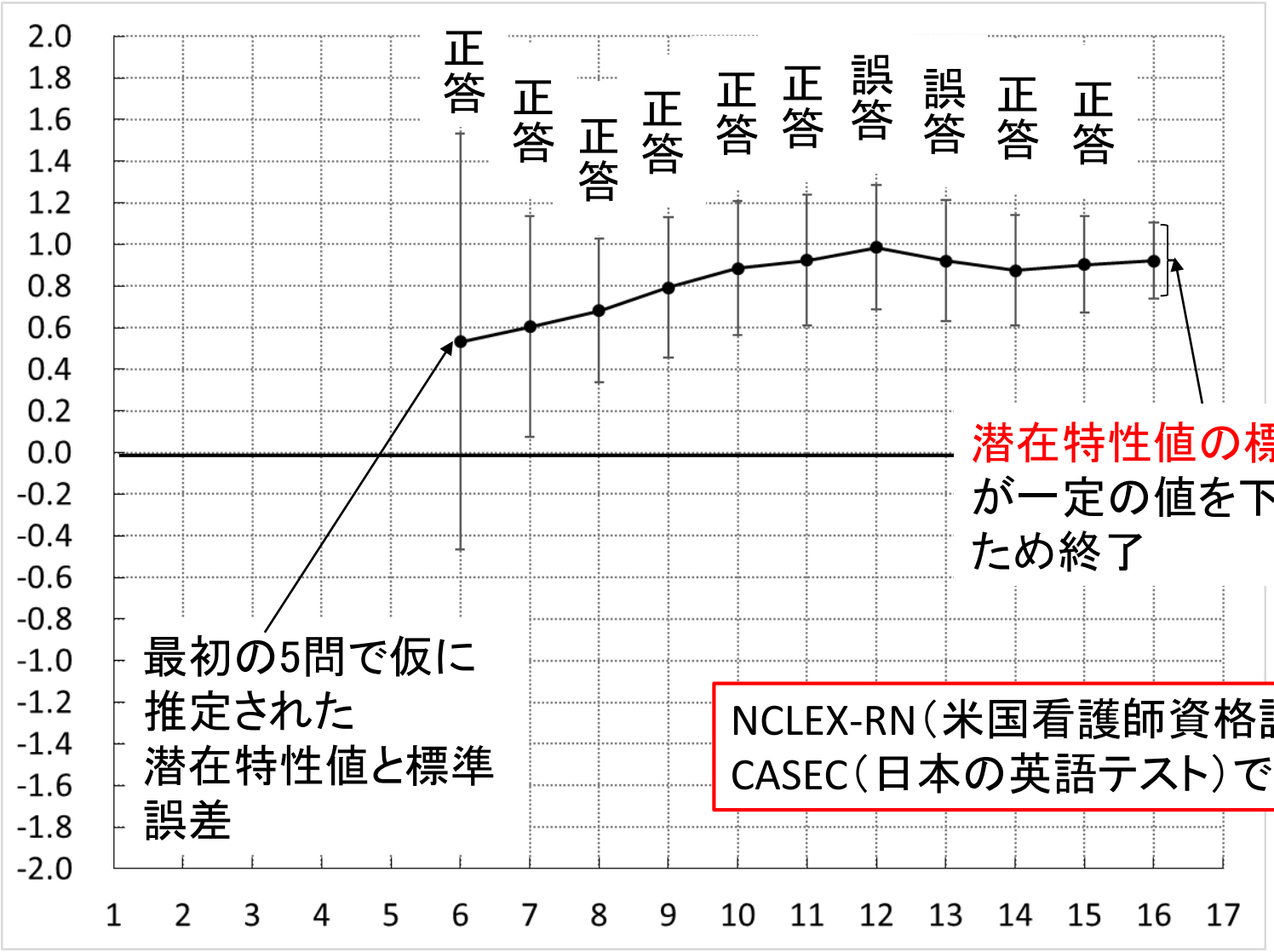


困難度が0以上の項目で、識別力が低いテスト版（2PL）

問	1	2	……	37	38	39	……	74	75
a_j	1.0	1.0	……	1.0	0.5	0.5	……	0.5	0.5
b_j	-3.7	-3.6	……	-0.1	0.0	0.1	……	3.6	3.7



コンピュータ適応型テスト（Computer Adaptive Test, CAT）

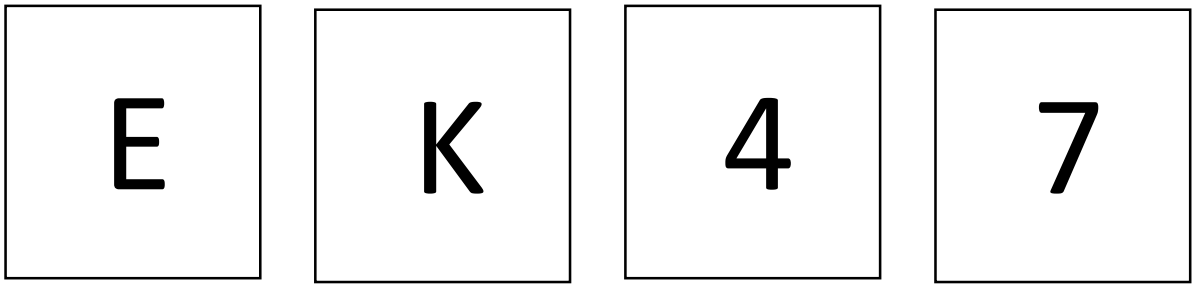


正解なら，少し難しめの項目で，
不正解なら，少し易しめの項目で，
} テスト情報関数が最大になる項目を次に提示。

- 項目反応理論(IRT)を用いる理由
- **IRTの概要**
  - IRTによる尺度化
  - 母数の推定方法
  - 母数の推定精度とテスト情報関数
  - **実践例**
  - 項目反応モデルの拡張
- IRTによる等化の概要
  - 等化の必要性
  - 等化の方法

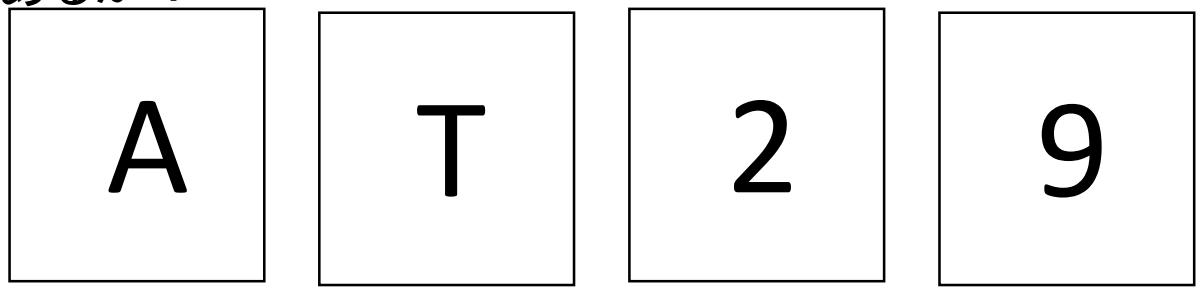
実践例1：Wason課題の項目母数推定（中村，2009）

「片面が母音ならば，もう片面は偶数である」ことを検証するためには，どれを裏返す必要があるか？



直接法（偶数）

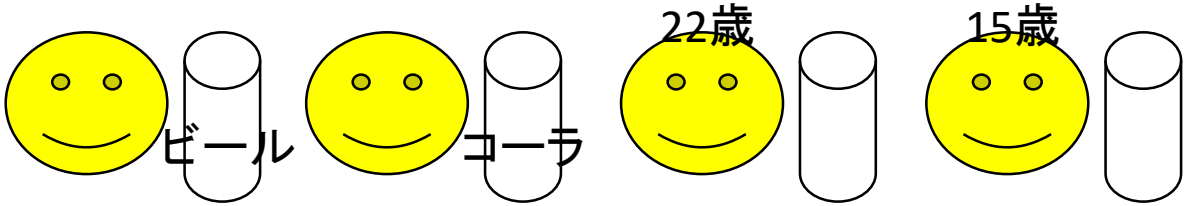
「片面がAならば，もう片面は2である」ことを検証するためには，どれを裏返す必要があるか？



直接法（A）

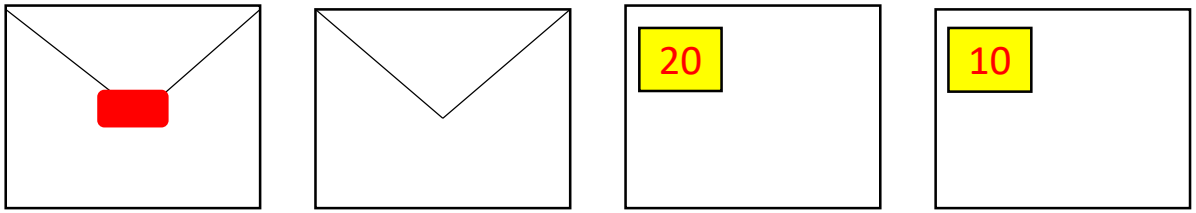
Wason課題の項目母数推定（中村，2009）

「ビールを飲んでいる人は，20歳以上でなければならない」ことを検証するためには，どの人に声かけすべきか？



ビール

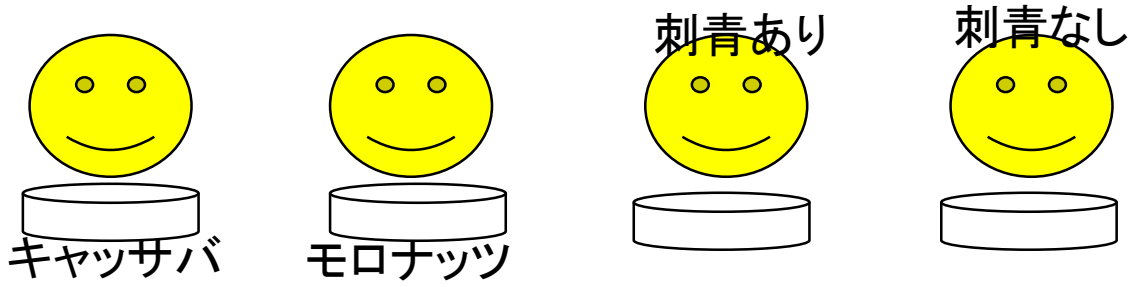
「郵便物に封がしてあるなら，表に20ドル切手が貼ってある」ことを検証するためには，どれを裏返す必要があるか？



郵便

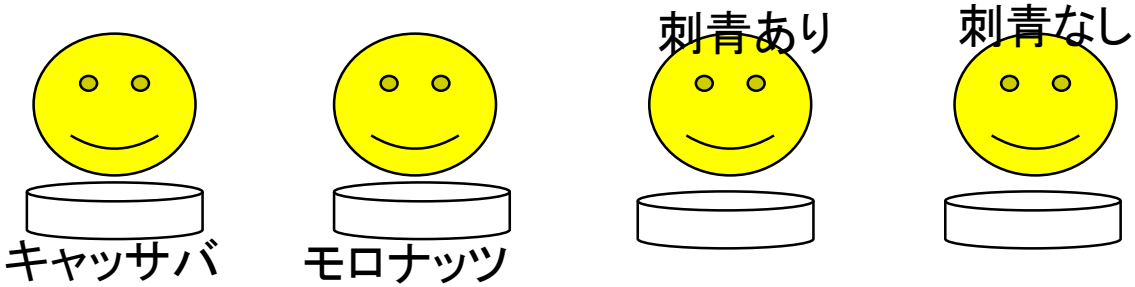
Wason課題の項目母数推定（中村，2009）

「部族の掟で，ある男がキャッサバを食べるならば，顔に刺青がなければならない．この規則を破っているかどうかを確かめることが課題」



キャッサバ課題
規則が「社会的交換」を含んでいる場合

「部族が常食するのはキャッサバまたはモロナッツなのだが，どちらかを確かめたい．この部族の掟で，キャッサバを食べるならば顔に刺青がなければならない」



キャッサバ課題
規則が「社会的交換」を含んでいない場合

- 6項目，323名から正誤データを収集
- 2PLによる項目反応モデルの当てはめ

Wason課題の項目母数推定（中村，2009）

- 既存の研究では，受験者の能力分布に依存した困難度指標しか得られなかった
- 項目単位での識別力を仮定することで，より細かい項目特性を明らかにしたい

→ 項目反応理論の適用

- 探索的カテゴリカル因子分析の結果，一因子性が確認された

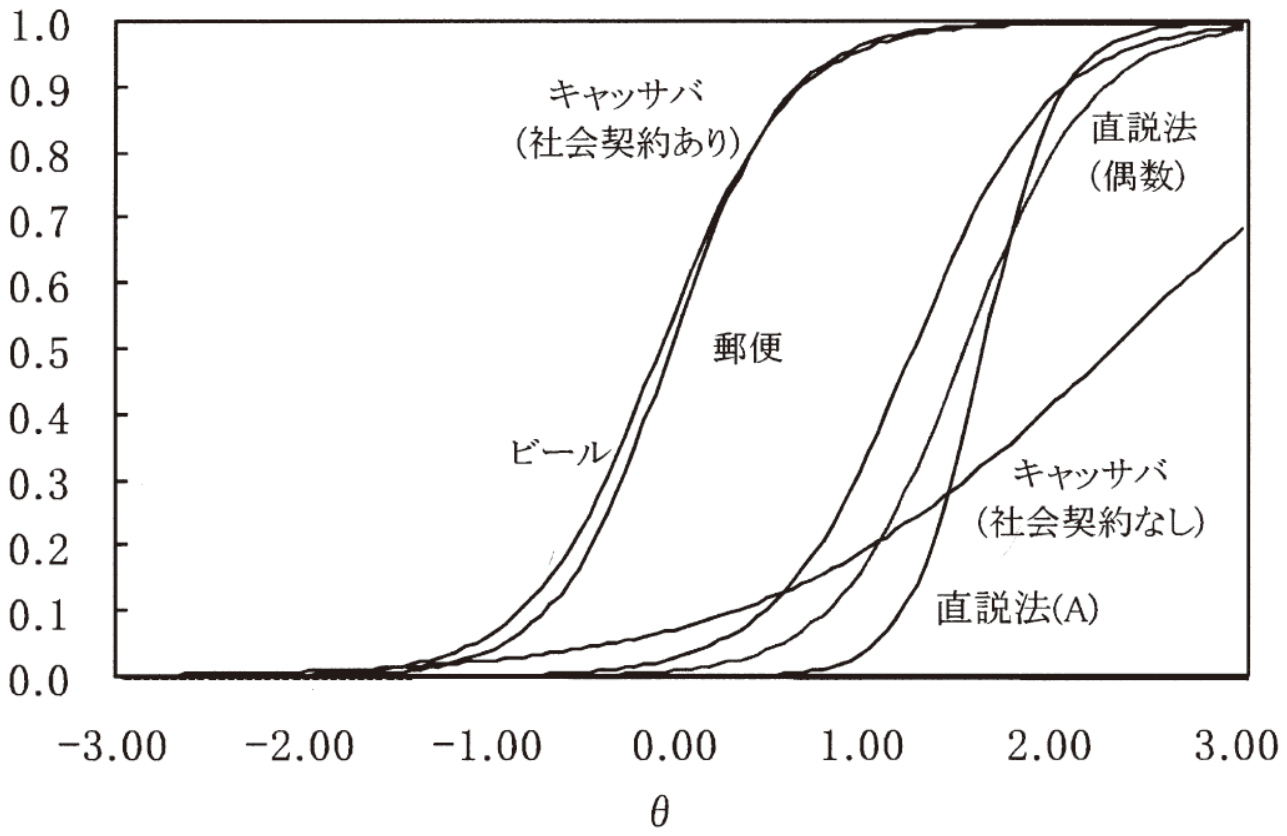
- 1PLと2PLのモデルの当てはまりの良さを比較したところ，2PLのほうが良かった（AICの値が小さいほうがモデルとデータの当てはまりが良いことを示す）。

AICによるモデル適合度の評価結果

1PL	2PL
1463.10	1442.91

Wason課題の項目母数推定（中村，2009）

課題名	a_j	b_j
直接法（偶数）	1.56	1.57
直接法（A）	2.22	1.70
ビール	1.79	-0.07
郵便	1.50	1.27
キャッサバ（あり）	1.67	-0.13
キャッサバ（なし）	0.67	2.20



項目識別力a_jの違いによって
項目の特徴が異なることが裏付けられた

実践例2：Big Five尺度短縮版の作成（並川ほか，2012）

- Big Five尺度：5因子×12項目＝60項目
- 短縮版の作成の根拠として，項目反応理論による項目母数及びテスト情報関数を用いた。
- 調査参加者（学生；$N = 2099$）に対して60項目に回答させ，多値型モデル（後述）により，5件法データを分析。

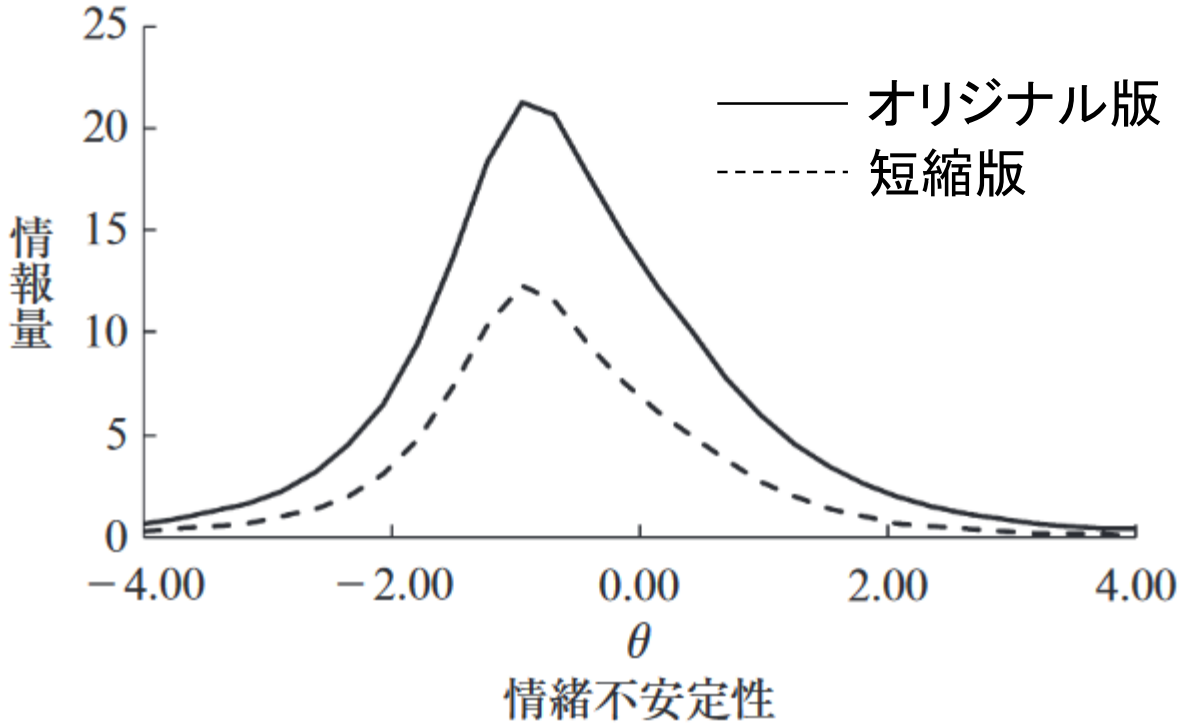
- 得られた項目母数から，識別力が高い項目で，かつ困難度（locationパラメタ）が互いに大きく異なる値となるような項目を選出し，平均テスト情報量（※）が一定の値を超えるような項目群を作成 →「短縮版」とした。

※ 平均テスト情報量：テスト情報関数に，重みとして標準正規分布の密度関数をかけ，θで積分したもの

実践例2・Big Five尺度短縮版の作成（並川ほか，2012）

- 作成された短縮版：「外向性」「情緒不安定性」5項目，
  「開放性」「調和性」6項目，
  「誠実性」7項目，合計 29 項目
- 短縮版データによる因子分析の結果も，5因子構造を支持。

- 短縮版において情報量の減少があるものの，一定の情報量は確保されている＝短縮版においても精度よく推定することが可能



IRTによる尺度化（まとめ）

- IRTにおける項目反応モデルは，カテゴリカルデータを用いた一次元の因子分析モデルと同等である。

- 2PL，3PLといったモデルを当てはめることにより，項目ごとの識別力を推定し，項目単位での測定精度の指標を得ることができる。
  - モデルの当てはまりの良さを評価することもできる。

- 潜在特性値ごとのテストの精度は，テスト情報関数によってあらわすことができる。また，潜在特性値の標準誤差を求めることで，潜在特性値の信頼区間を推定することができる。

- テスト情報関数を手がかりに，特定の潜在特性値の受験者層に精度が高くなるような項目群を選抜することができる。

- 項目反応理論(IRT)を用いる理由
- IRTの概要
  - テストで得られるデータ
  - IRTによる尺度化
  - 母数の推定方法
  - 母数の推定精度とテスト情報関数
  - 実践例
  - 項目反応モデルの拡張
- IRTによる等化の概要
  - 等化の必要性
  - 等化の方法

二値型・多値型・多相モデル

- 二値型モデル
  - 1パラメタ・ロジスティックモデル
  - 2パラメタ・ロジスティックモデル
  - 3パラメタ・ロジスティックモデル

- 多値型モデル
  - 段階反応モデル
  - 部分採点モデル
  - 名義反応モデル

- ラッシュモデル（Rasch model）
  - 多相ラッシュモデル（Many-facet Rasch model）

- 多母集団モデル

段階反応モデル（graded response model, Samejima, 1969）

能力と「あるカテゴリを選ぶ確率」を、以下のように記述する。

例・4カテゴリ（4件法）の場合

カテゴリ「1」または「2」または「3」または「4」と反応する確率：$P_{j1}^{+}(\theta) = 1$

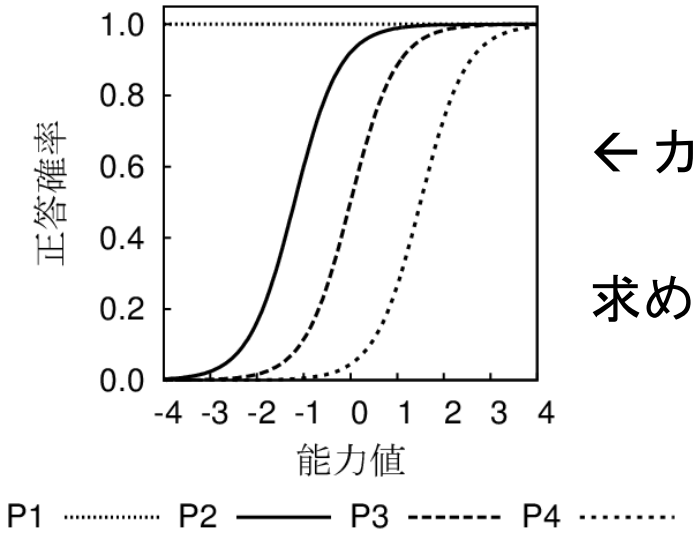
カテゴリ「2」または「3」または「4」と反応する確率：$P_{j2}^{+}(\theta)$

カテゴリ「3」または「4」と反応する確率：$P_{j3}^{+}(\theta)$

カテゴリ「4」と反応する確率：$P_{j4}^{+}(\theta)$

いずれのカテゴリでもない確率：$P_{j5}^{+}(\theta) = 0$

$$P_{jk}^{+}(\theta) = \frac{1}{1 + \exp(-1.7a_j(\theta - b_{jk}))} \ (k = \{2,3,4\})$$



← カテゴリ境界特性曲線

求めるべき母数：$a_j, b_{j2}, b_{j3}, b_{j4}$

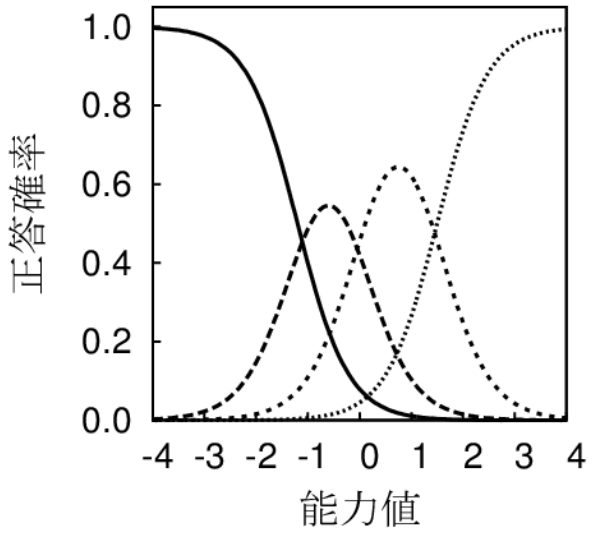
カテゴリ「１」と反応する確率は、
カテゴリ「１」または「２」または「３」または「４」と反応する確率：$P_{j1}^{+}(\theta)$から
カテゴリ「２」または「３」または「４」と反応する確率：$P_{j2}^{+}(\theta)$を引けば求まる

$$P_{j1}(\theta) = P_{j1}^{+}(\theta) - P_{j2}^{+}(\theta) = 1 - P_{j2}^{+}(\theta)$$

カテゴリ「２」と反応する確率は、
カテゴリ「２」または「３」または「４」と反応する確率：$P_{j2}^{+}(\theta)$から
カテゴリ「３」または「４」と反応する確率：$P_{j3}^{+}(\theta)$を引けば求まる。

$$P_{j2}(\theta) = P_{j2}^{+}(\theta) - P_{j3}^{+}(\theta)$$

・・・・・・以降，隣り合ったカテゴリについて，差をとると，



← 項目反応カテゴリ特性曲線　を得る。

ある能力をもつ受験者が，あるカテゴリで
あるとされる確率を，カテゴリ別に示す。

- 項目反応理論(IRT)を用いる理由
- IRTの概要
  - IRTによる尺度化
  - 母数の推定方法
  - 母数の推定精度とテスト情報関数
  - 実践例
  - 項目反応モデルの拡張
- IRTによる等化の概要
  - 等化の必要性
  - 等化の方法

【等化】
あるテストにおいて同一設計のもとで作成された異なる版の得点を，共通尺度上の得点に変換する統計的な手続きのこと．この手続きはこれらの版がすべて同じ特性を測定していることを前提とする（後略）．

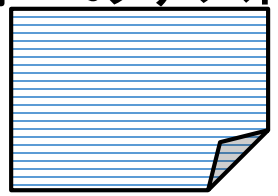
【リンク】（リンキング）
複数のテスト得点を同じひとつの尺度上の得点として表現し，互いに比較可能なようにする手続きの総称．共通尺度である必要は必ずしもない．等化は一番強い仮定のもとで行われるリンキングの一つ．**→【等化】**

「テスト・スタンダード」（日本テスト学会・編，p.220, p.224）より

等化の必要性・・・ケース1

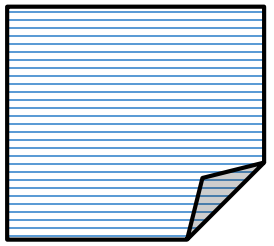
最初の4項目が
同一のテスト版

集団A
規準集団



項目
母数
A

集団B



項目
母数
B

集団における平均的能力の違いによって，推定される困難度が変わる

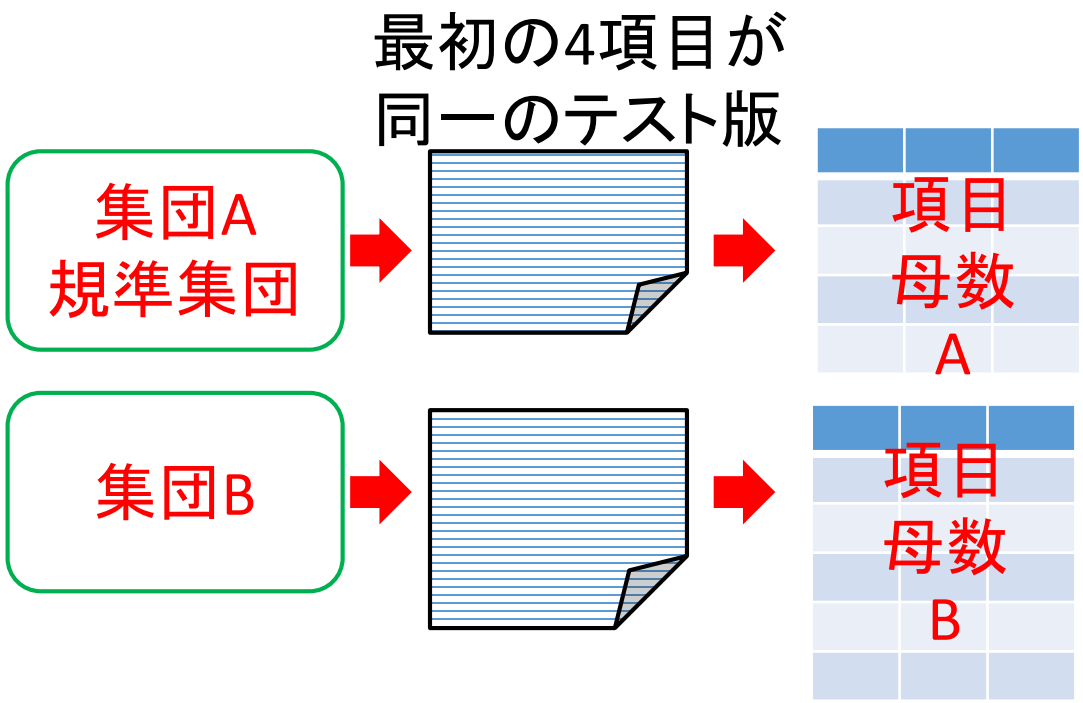
集団Bの困難度を2倍することで，集団Aの困難度に合わせることができる

集団Bだけに出題された項目5についても，同様に2倍すればよい

集団A	識別力	困難度
項目1	0.8	−1.2
項目2	1.2	−0.8
項目3	0.8	0.4
項目4	1.0	1.2

集団B	識別力	困難度
項目1	0.8	−0.6
項目2	1.2	−0.4
項目3	0.8	0.2
項目4	1.0	0.6
項目5	1.0	0.8

等化の必要性･･･ケース2



集団A	識別力	困難度
項目1	0.8	−1.2
項目2	1.2	−0.8
項目3	0.8	0.4
項目4	1.0	1.2

集団B	識別力	困難度
項目1	0.8	−2.2
項目2	1.2	−1.8
項目3	0.8	−1.4
項目4	1.0	0.2
項目5	1.0	0.8

集団Aの尺度で表現された集団Bの困難度は，集団Bから得られたもとの困難度に1を足すことで得られる
項目5も同様に変換することによって，仮にこの項目を集団Aで出題した際の困難度を推定することが可能

集団Bの方が集団Aより困難度が1だけ低い
＝集団Bの方が集団Aより能力が高い

等化の必要性

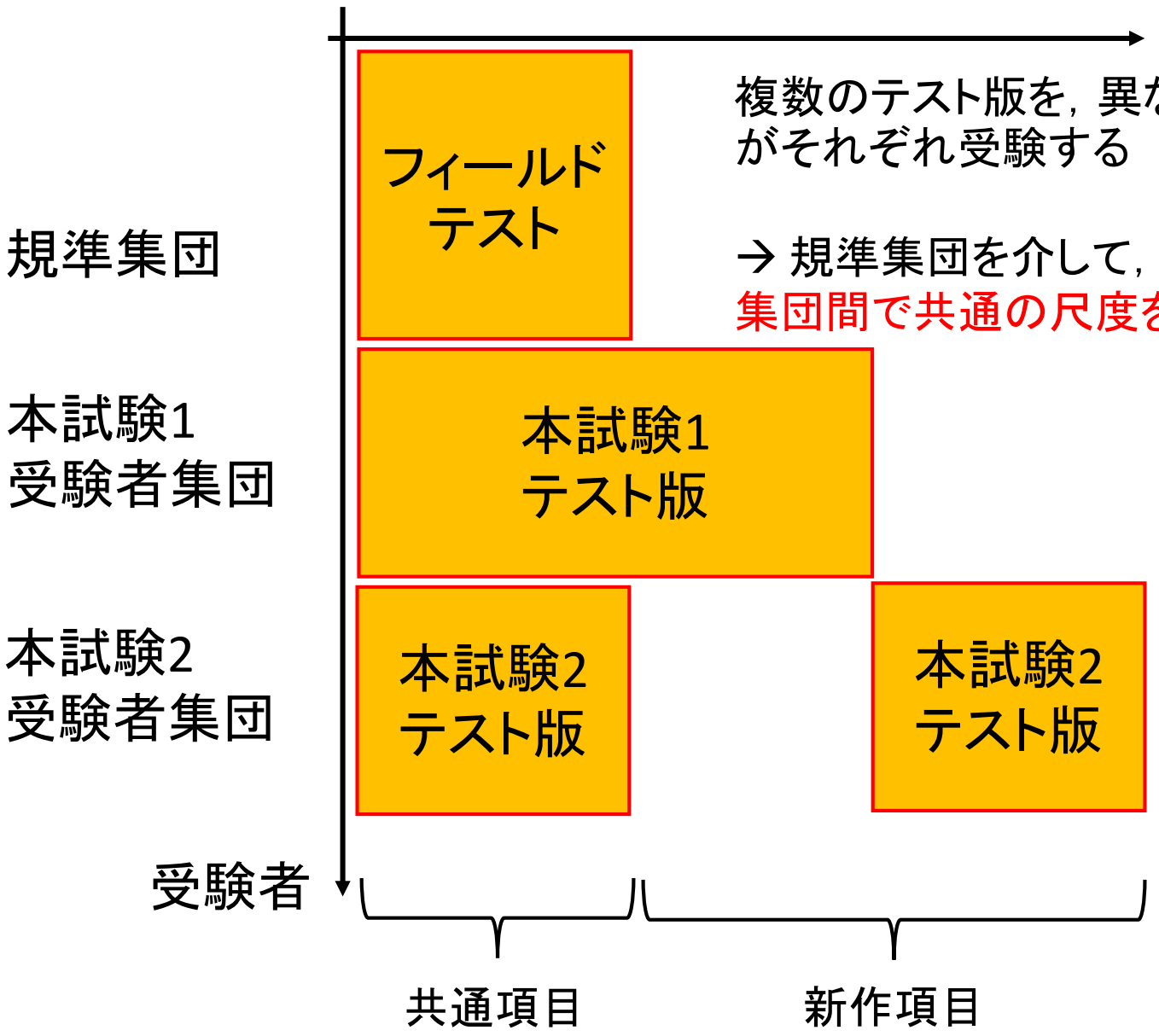
集団A	識別力	困難度
項目1	0.8	−1.2
項目2	1.2	−0.8
項目3	0.8	0.4
項目4	1.0	1.2

集団B	識別力	困難度
項目1	0.8	−2.2
項目2	1.2	−1.8
項目3	0.8	−1.4
項目4	1.0	0.2
項目5	1.0	0.8

- 項目母数が異なるのは，受験した集団の間で，能力が異なるため
- 二つのテストで共通の尺度で項目母数を表現する手続きにより，一方の集団の項目母数を，他方の集団の尺度上で表現できる → 等化
- TOEIC，TOEFLなどの大規模テストでは，多数の受験会場間で受験者集団の能力分布が異なる → 等化による共通尺度化により，どの受験会場でも共通の意味をもつスコアを算出することができる

- 項目反応理論(IRT)を用いる理由
- IRTの概要
  - IRTによる尺度化
  - 母数の推定方法
  - 母数の推定精度とテスト情報関数
  - 実践例
  - 項目反応モデルの拡張
- IRTによる等化の概要
  - 等化の必要性
  - 等化の方法

規準となる集団を定義し，その尺度上でテストスコアを表示するために



複数のテスト版を，異なる受験者集団がそれぞれ受験する

→ 規準集団を介して，複数の受験者集団間で共通の尺度を求めたい。

IRTに基づく項目母数推定値を用いた等化
（項目母数の尺度変換）

テスト版Aの尺度上で表現された，
テスト版Bの項目母数の変換後の値：a_{jB}^{*}（識別力） b_{jB}^{*}（困難度）

テスト版Bのサンプルから得られた，
項目母数（変換前の値）：a_{jB}（識別力） b_{jB}（困難度）

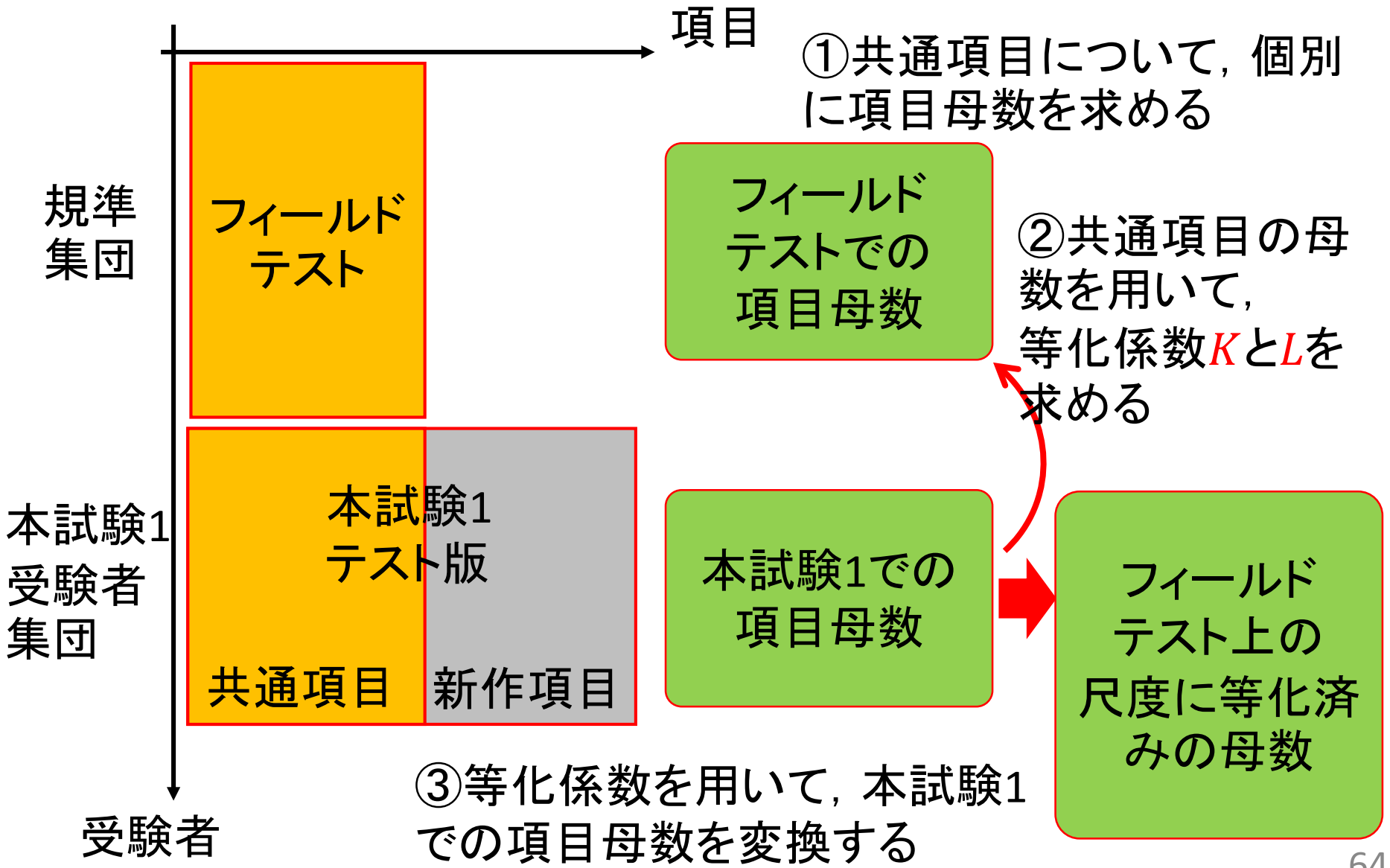
$$a_{jB}^{*} = a_{jB}/K$$
$$b_{jB}^{*} = Kb_{jB} + L$$

K倍してLを足す，という操作で，
テスト版Bの受験者に提示した項目の母数が，
テスト版Aの受験者に提示した場合での項目母数に変換できる．

ただし，二つのテスト版には共通項目が存在することが前提．

※ ICCは不変なため，このような線形変換が可能。

個別推定による等化


項目
規準集団
フィールドテスト
本試験1受験者集団
本試験1テスト版
共通項目
新作項目
受験者
①共通項目について，個別に項目母数を求める
フィールドテストでの項目母数
②共通項目の母数を用いて，等化係数KとLを求める
本試験1での項目母数
フィールドテスト上の尺度に等化済みの母数
③等化係数を用いて，本試験1での項目母数を変換する

等化係数KとLの求め方

- Mean & Sigma法（Marco, 1977）
  - 困難度のみを用いた方法
- Mean & Mean法（Loyd & Hoover, 1980）
  - 識別力を加味した方法

- Haebara法（Haebara, 1980）
- Stocking-Lord法（Stocking & Lord, 1983）
  - 規準集団から得られたICCに，他のテスト版から得られたICCを近づける方法

以上の方法は，一度に1つのテスト版しか規準集団に等化できない

- Calr法（前川，1991; Arai and Mayekawa, 2011）
  - 複数のテスト版を，同時に規準集団へ等化できる。

項目バンクの構築とその運用

- 項目特性を推定するために「フィールドテスト」を行っておく。
- フィールドテストで得られた正誤データを分析し，項目特性を推定する。
- 推定した項目特性を，項目の本文，選択枝とともにデータベース化しておく（項目バンク）。
- 「本試験」では，あらかじめターゲットとする潜在特性値の範囲を決めておき，その範囲で情報量が大きくなるような項目群を項目バンクから選択する。→ 本試験を複数回行う過程で，項目を使いまわすこともありうる

- 項目バンクを用いた大規模テストを行う場合，公平なテスト実施のため，項目の内容はテスト実施後も秘匿する必要がある。
- しかし，日本のテストは，テスト項目が実施後にすべて公開されるものが多い。すべて公開されると，項目バンクの意味がなくなってしまう。

IRTによる等化の概要（まとめ）

- 複数の集団から得られた項目特性や潜在特性値を比較するために，複数の尺度をまたいで定義される共通尺度を定め，共通尺度に他の尺度を合わせる操作を行う。この操作を「等化」という。

- 潜在特性値が複数のテスト間で共通の意味をもつようにするためには，適切なテストデザイン及び等化の計画をたて，それらが誤りなく行われなければならない。また，それぞれの尺度が同一の概念を測定している必要がある。

- 項目母数が既知の項目を多数用意して項目バンクに登録し，受験者に提示する項目をテスト実施者がコントロールすることで，受験者の能力の推定により適した項目を用いたテストを行うことができる。等化は，項目バンク構築のために必要不可欠である。

- 日本で一般的な「テスト実施後，項目をすべて公開する」タイプのテストでは，項目バンクを用いた大規模テストを実施するのは困難。

参考文献

Arai, S., Mayekawa, S.（2011）. A comparison of equating methods and linking designs for developing an item pool under item response theory. *Behaviormetrika*, **38（1）**, 1-16.

Haebara, T.（1980）. Equating logistic ability scales by a weighted least squares method. *Japanese Psychological Research*, **22（3）**, 144-149.

加藤健太郎・山田剛史・川端一光（2014）. Rによる項目反応理論 オーム社

Loyd, B.H., & Hoover, H.D.（1980）. Vertical equating using the Rasch Model. *Journal of Educational Measurement*, **17**, 179-193.

前川眞一（1991）. パラメタの推定 芝祐順［編］ 項目反応理論―基礎と応用 東京大学出版会

Marco, G.L.（1977）. Item characteristic curve solutions to three intractable testing problems. *Journal of Educational Measurement*, **14**, 139-160.

中村國則（2009）. 項目反応理論を用いた4枚カード問題の分析 心理学研究, **80(5)**, 436-441.

並川努・谷伊織・脇田貴文・熊谷龍一・中根愛・野口裕之（2012）. Big Five尺度短縮版の開発と信頼性と妥当性の検討 心理学研究, **83(2)**, 91-99.

日本テスト学会［編］（2007）. テスト・スタンダード―日本のテストの将来に向けて 金子書房

Samejima, F.（1969）. Estimation of latent ability using a response pattern of graded scores. *Psychometric Monograph*, **17**, Psychometric Society.

Stocking, M.L., & Lord, F.M.（1983）. Developing a common metric in item response theory. *Applied Psychological Measurement*, **7**, 201-210.

Rを用いた項目反応理論による分析方法・実践例の紹介

登藤　直弥

（筑波大学人間系）

IRTに基づく分析を行うためのソフトウェアについて

本チュートリアルセミナーで使用するソフトウェア

- 項目反応理論（Item Response Theory, IRT）に基づく分析
  - 様々なソフトウェアによって実行可能
    - BILOG-MG（Zimowski, Muraki, Mislevy, & Bock, 1996）
    - EasyEstimation（熊谷，2009）
    - Exametrika（Shojima，2014）
- R（R Core Team, 2017）
  - 統計解析用のソフトウェア
  - 心理学領域の研究で用いられることも多くなってきた
  - インストールさえできれば，誰でもすぐに利用可能
  - IRTに基づく分析も実行可能
- 本発表
  - IRTによる項目母数と潜在特性値の推定ならびに等化などをRを用いて実行する方法を紹介

テストデータの解析例

テストデータの例

- ltmパッケージの中に入っている実データを使用
  - The Law School Admission Testへの解答結果を採点
    - 受験者数：1000人
    - 項目数：Section IVに含まれる5項目
    - 正答なら1，誤答なら0と入力
  - 以下の変数が含まれる
    - ID: 受験者のID
    - Item1: １番目の項目への解答結果
    - Item2: ２番目の項目への解答結果
    - Item3: ３番目の項目への解答結果
    - Item4: ４番目の項目への解答結果
    - Item5: ５番目の項目への解答結果
    - SS（独自に追加）: Item1-5の合計（素点）
    - class（独自に追加）：素点に基づく受験者のクラス分け
      - 1-5で表され，値が高くなるほど，素点の高いクラスになる

  ⇒ このデータをIRTに基づき分析し，項目やテストの特性を評価する
- ltmパッケージを読み込んだ（library(ltm)した）後
  - R Consoleに”data(LSAT)”と入力しても入手可能
    - data( )：指定された（括弧の中にある）データセットをRに読み込む

テストデータの例

ID	Item1	Item2	Item3	Item4	Item5	SS	class
1	0	0	0	0	0	0	1
2	0	0	0	0	0	0	1
3	0	0	0	0	0	0	1
4	0	0	0	0	1	1	1
5	0	0	0	0	1	1	1
6	0	0	0	0	1	1	1
7	0	0	0	0	1	1	1
8	0	0	0	0	1	1	1
9	0	0	0	0	1	1	1
10	0	0	0	1	0	1	1
11	0	0	0	1	0	1	1

Rにテストデータを読み込む方法

- Rにテストデータ（CSVファイル）を読み込む
  - read.csv関数（CSVファイルを読み込む）を利用して読み込み
- read.csv関数の使い方（今の場合）
  - read.csv(file = "LSAT.csv", header = TRUE, row.names = 1)
    - file: テストデータの入っているCSVファイルの名前を指定する引数
      - 今の場合，利用するデータは"LSAT.csv"に入っているので，これを指定する
    - header: CSVファイルの１行目に変数名が記載されているか否か
      - ”TRUE（はい）”と指定することでCSVファイルの１行目が変数名と認識される
    - row.names: 受験者IDが何列目に記載されているかを指定する引数
      - ”1”と指定することでCSVファイルの１列目が受験者IDと認識される
  - TEST <- read.csv(file = "LSAT.csv", header = TRUE, row.names = 1)

⇒　読み込んだCSVファイルのテスト（TEST）データがTESTに

R Consoleに“TEST”と打ち込んでEnter
⇒　読み込んだデータが表示される

```
> TEST
   Item1 Item2 Item3 Item4 Item5 SS class
1      0     0     0     0     0  0     1
2      0     0     0     0     0  0     1
3      0     0     0     0     0  0     1
4      0     0     0     0     1  1     1
5      0     0     0     0     1  1     1
6      0     0     0     0     1  1     1
7      0     0     0     0     1  1     1
8      0     0     0     0     1  1     1
9      0     0     0     0     1  1     1
10     0     0     0     1     0  1     1
11     0     0     0     1     0  1     1
```

項目特性の把握

正答率の計算方法

- 光永先生の資料より
  - 古典的テスト理論では正答率を用いて項目の特性を把握
- Rでの正答率の計算
  - colMeans関数（列毎に平均を計算する）で計算可能

  ⇒　1とOの二値データの平均は正答者の割合になる
- colMeans関数の使い方（今の場合）
  - colMeans(x = TEST[, 1:5], na.rm = TRUE)
    - x: テストデータを指定する引数
      - テストデータは行が受験者，列が項目を表すようにする
      - TEST[, 1:5]とすることで，Item1-Item5のデータのみが対象に
    - na.rm: 欠測値（NA）がある場合に，計算に含めないか
      - 普通はNAは計算時に除外するので”TRUE（はい）”とする
  - CRR <- colMeans(x = TEST[, 1:5], na.rm = TRUE)

  ⇒　正答率（Correct Response Rate）の計算結果がCRRに

R Consoleに“CRR”と打ち込んでEnter
⇒　計算結果が表示される

```
> CRR <- colMeans(x = TEST, na.rm = TRUE)
> CRR
Item 1 Item 2 Item 3 Item 4 Item 5
 0.924  0.709  0.553  0.763  0.870
>
```

易しい項目が多い

I-T相関の計算方法

- 光永先生の資料より
  - 古典的テスト理論ではI-T相関で項目の特性を把握
- RでのI-T相関の計算
  - cor関数（相関を計算する）で素点と各項目得点との相関を計算
- cor関数の使い方（今の場合）
  - cor(x = TEST[, 1:5], y = TEST[, 6], use = "pairwise.complete.obs")
    - x: 項目得点（テストデータ）を指定する引数
    - y: 素点を指定する引数
      - TEST[, 6]とすることで，SSのデータが対象に
    - use: 欠測値（NA）がある場合の相関の計算方法を指定する引数
      - "pairwise.complete.obs"とすると，素点と各項目得点との相関を計算する際に，NAのある受験者のデータを除いて計算を行う
  - ITcor<-cor(x=TEST[,1:5],y=TEST[,6],use="pairwise.complete.obs")

⇒ I-T相関（I-T correlation）の計算結果がITcorに

R Consoleに“ITcor”と打ち込んでEnter
⇒　計算結果が表示される

```
> ITcor <- cor(x = TEST, y = SS, use = "pairwise.complete.obs")
> ITcor
             [,1]
Item 1 0.3620104
Item 2 0.5667721
Item 3 0.6184398
Item 4 0.5344183
Item 5 0.4353664
```

- I-T相関の高い項目が多い

⇒ 除外すべき項目はみられない

１次元性の検討

１次元性の仮定とその検討方法

- 光永先生の資料より
  - IRTに基づきテストデータや質問紙への回答を分析する
  ⇒　１次元性が仮定されている
- IRTのモデル
  - 因子分析のモデルと関係が深い
  ⇒　１次元性の仮定の検討
  ⇒　因子分析で利用される方法を用いて検討することが多い
    - スクリープロット
    - 平行分析（Horn, 1965）
    ⇒　Rでの実行方法については質問紙データの解析例にて解説
- 上記の方法を用いてテストデータ例の次元数を検討
  - １次元性が示唆される

項目反応モデルの項目母数の推定

本チュートリアルセミナーで扱う項目反応モデル

- 本チュートリアルセミナー
  - 以下の項目反応モデルを扱う
    - 1パラメタ・ロジスティックモデル
      - 1 Parameter Logistic Modelなので，1PLMと略す
    - 2パラメタ・ロジスティックモデル
      - 2 Parameter Logistic Modelなので，2PLMと略す
    - 3パラメタ・ロジスティックモデル
      - 3 Parameter Logistic Modelなので，3PLMと略す
    - 段階反応モデル
      - Graded Response Modelなので，GRMと略す

      ⇒ のちほど，質問紙データの解析例において解説
- 項目母数の推定
  - 潜在特性値は平均が0，分散が1の正規分布（標準正規分布）に従うと仮定して推定を行うのが一般的

Rで1PLMの項目母数を求める方法

- 光永先生の資料より
  - IRTにおいては項目母数の値よりその項目の特性を把握
- Rで1PLMを仮定したときの項目母数を周辺最尤推定する
  - est関数（項目母数を推定する）を利用して項目母数を推定

  ⇒　est関数はirtoysパッケージの中に入っている関数
- est関数の使い方（今の場合）
  - est(resp = TEST[, 1:5], model = "1PL", engine = "ltm")
    - resp: テストデータを指定する引数
    - model: データに対してどのようなモデルを当てはめるか指定する引数
      - "1PL"と指定することでデータに対し1PLMが仮定される
    - engine: 何を利用して項目母数の推定をするかを指定する引数
      - "ltm"と指定することでltmパッケージを利用して項目母数を推定
  - para.1PL<-est(resp=TEST[,1:5],model="1PL",engine="ltm")

  ⇒　項目母数の推定結果がpara.1PLに

R Consoleに”para.1PL”と打ち込んでEnter
⇒　推定結果が表示される

```
>    para.1PL
$est
          [,1]        [,2] [,3]
Item1 0.7551354 -3.6152642    0
Item2 0.7551354 -1.3224200    0
Item3 0.7551354 -0.3176311    0
Item4 0.7551354 -1.7300895    0
Item5 0.7551354 -2.7801707    0
```

- $estに項目母数の推定値
  - □１列目：1.7×識別力母数
  - □２列目：困難度母数
  - □３列目：当て推量母数

- 推定値からわかること
  - 識別力が項目間で同一
  - 易しい項目が多い

Rで2PLMの項目母数を求める方法

- 光永先生の資料より
  - IRTにおいては項目母数の値よりその項目の特性を把握
- Rで2PLMを仮定したときの項目母数を周辺最尤推定する
  - est関数（項目母数を推定する）を利用して項目母数を推定

  ⇒　est関数はirtoysパッケージの中に入っている関数
- est関数の使い方（今の場合）
  - est(resp = TEST[, 1:5], model = "2PL", engine = "ltm")
    - resp: テストデータを指定する引数
    - model: データに対してどのようなモデルを当てはめるか指定する引数
      - "2PL"と指定することでデータに対し2PLMが仮定される
    - engine: 何を利用して項目母数の推定をするかを指定する引数
      - "ltm"と指定することでltmパッケージを利用して項目母数を推定
  - para.2PL<-est(resp=TEST[,1:5],model="2PL",engine="ltm")

  ⇒　項目母数の推定結果がpara.2PLに

R Consoleに”para.2PL”と打ち込んでEnter
⇒　推定結果が表示される

```
>   para.2PL
$est
          [,1]        [,2] [,3]
Item1 0.8253717 -3.3597333    0
Item2 0.7229498 -1.3696501    0
Item3 0.8904752 -0.2798981    0
Item4 0.6885500 -1.8659193    0
Item5 0.6574511 -3.1235746    0
```

- $estに項目母数の推定値
  - １列目：　1.7×識別力母数
  - ２列目：困難度母数
  - ３列目：当て推量母数

- 推定値からわかること
  - 識別力が項目毎に異なる
  - 易しい項目が多い

Rで3PLMの項目母数を求める方法

- 光永先生の資料より
  - IRTにおいては項目母数の値よりその項目の特性を把握
- Rで3PLMを仮定したときの項目母数を周辺最尤推定する
  - est関数（項目母数を推定する）を利用して項目母数を推定

  ⇒　est関数はirtoysパッケージの中に入っている関数
- est関数の使い方（今の場合）
  - est(resp = TEST[, 1:5], model = "3PL", engine = "ltm")
    - resp: テストデータを指定する引数
    - model: データに対してどのようなモデルを当てはめるか指定する引数
      - "3PL"と指定することでデータに対し3PLMが仮定される
    - engine: 何を利用して項目母数の推定をするかを指定する引数
      - "ltm"と指定することでltmパッケージを利用して項目母数を推定
  - para.3PL<-est(resp=TEST[,1:5],model="3PL",engine="ltm")

  ⇒　項目母数の推定結果がpara.3PLに

R Consoleに”para.3PL”と打ち込んでEnter
⇒　推定結果が表示される

```
> para.3PL
$est
            Dscrmn     Dffclt     Gussng
Item1   0.2945829 -8.2376044 0.09306093
Item2  12.0612683  0.9892132 0.64301268
Item3  14.6898035  1.0033451 0.45379015
Item4  20.8485498 -0.5737457 0.01076137
Item5   0.4268108 -3.9246143 0.21505474
```

- 3PLMを利用する場合，安定した解が得られないことも

- $estに項目母数の推定値
  - １列目：　1.7×識別力母数
  - ２列目：困難度母数
  - ３列目：当て推量母数

R Consoleに”para.3PL”と打ち込んでEnter
⇒ “$se”に項目母数の標準誤差が記載

```
> para.3PL
$est
           Dscrmn     Dffclt     Gussng
Item1   0.2945829 -8.2376044 0.09306093
Item2  12.0612683  0.9892132 0.64301268
Item3  14.6898035  1.0033451 0.45379015
Item4  20.8485498 -0.5737457 0.01076137
Item5   0.4268108 -3.9246143 0.21505474
```

```
$se
      [,1] [,2] [,3]
[1,]   NaN   NaN   NaN
[2,]   NaN   NaN   NaN
[3,]   NaN   NaN   NaN
[4,]   NaN   NaN   NaN
[5,]   NaN   NaN   NaN
```

- 適当な数値が入っていない
⇒ 推定が上手くいっていない
- 推定結果が得られる
⇒ 標準誤差を参照することが重要

項目母数の推定を行った論文例

Japanese Journal of Educational Psychology, 1991, 39, 204—213

資　料

項目反応理論に基づく計算課題の難易度の推定

喜　岡　恵　子*

ESTIMATION OF THE DIFFICULTY LEVELS OF CALCULATION TASKS BASED ON ITEM RESPONSE THEORY

Keiko KIOKA

The purpose of this research was to estimate the difficulty levels of tasks in arithmetic calculations in order to investigate the optimum arrangement of the tasks for teaching. The six forms of tests were constructed so that each form should include items of the calculation tasks for each grade at the elementary school level, and they were administered to 2822 pupils. By means of the two-parameter logistic model, the item parameters were estimated for the 221 items contained in the six forms of tests. The rusults showed that the model was satisfactorily fitted to the data. Furthermore, inappropriateness of conventional order of calculation tasks in teaching was also discussed.

Key words : item response theory, difficulty levels, arithmetic calculations, optimum arrangement for teaching.

喜岡 (1991)における項目母数の推定

- 小学生の計算能力を測定するためのテストを作成
  - 出題範囲は小学校の教育課程において学習する整数・小数・分数の四則演算に限定
  - 学習指導要領に準拠して小学1-6年生版の6つのテストを作成
  - 各項目は自由記述式で，正答の場合は1，それ以外の場合には0と採点された
- 調査対象者
  - 小学1年生から中学1年生までの児童・生徒2,822人
- IRTに基づく分析を行う際に用いた項目反応モデル
  - 2PLM

⇒　受験者集団に依らない項目の難易度が得られる

⇒　データに基づく指導順序の検討が可能に

喜岡 (1991)における推定結果（抜粋）

TABLE 2　各項目の識別力及び困難度（一部略）

項目番号	項目（計算課題）	識別力	識別力の推定誤差	困難度	困難度の推定誤差
(83)	7×0	0.2	2.17	−11.3	
(1)	4+3	1.3	0.67	−3.8	2.82
(4)	9+1	2.0	0.64	−3.1	0.46
(2)	8+0	2.0	0.60	−3.1	0.40
(14)	9−0	2.0	0.47	−2.9	0.25
(13)	5−1	1.9	0.43	−2.8	0.27
(15)	3−3	1.6	0.36	−2.8	0.34
(3)	6+7	1.6	0.22	−2.5	0.17
(28)	5+2+1	2.0	0.34	−2.4	0.13
(6)	50+6	2.0	0.33	−2.4	0.12
(16)	14−1	2.0	0.33	−2.4	0.12
(8)	2+80	2.0	0.32	−2.4	0.11
(18)	12−2	2.0	0.31	−2.3	0.11
(5)	32+4	2.0	0.30	−2.2	0.09
(20)	10−4	2.0	0.29	−2.2	0.09
(31)	8−2−1	1.7	0.27	−2.2	0.12
(29)	4+6+3	2.0	0.21	−2.1	0.08
(34)	10−5−4	2.0	0.28	−2.1	0.08
(9)	30+40	2.0	0.28	−2.1	0.08
(62)	173−95	1.4	0.32	−1.3	0.21
(68)	712−361	2.0	0.32	−1.3	0.10
(82)	7×7	2.0	0.23	−1.2	0.08
(63)	112−19	1.9	0.31	−1.2	0.10
(80)	5×9	2.0	0.31	−1.2	0.09
(71)	534−438	2.0	0.31	−1.2	0.08
(79)	8×7	2.0	0.30	−1.1	0.07
(77)	6×8	2.0	0.30	−1.1	0.07
(70)	922−229	1.6	0.19	−1.1	0.10
(72)	504−126	2.0	0.30	−1.0	0.06
(86)	27×3	2.0	0.31	−0.9	0.12
(107)	0.3+0.5	2.0	0.28	−0.8	0.09
(112)	0.7−0.6	2.0	0.28	−0.7	0.09
(108)	4.2+5.3	2.0	0.27	−0.7	0.08
(99)	80÷2	2.0	0.27	−0.6	0.08
(110)	4.4+1.8	2.0	0.27	−0.6	0.08
(113)	8.7−1.5	2.0	0.27	−0.6	0.08
(129)	60÷15	2.0	0.37	−0.6	0.16
(98)	99÷3	2.0	0.26	−0.6	0.07

潜在特性値の推定

Rで潜在特性値を求める方法

- 光永先生の資料より
  - IRTにおいては潜在特性の値よりその受験者の特性を把握
- Rで潜在特性値を推定する
  - mlebme関数を利用して潜在特性値を推定

  ⇒　mlebme関数はirtoysパッケージの中に入っている関数
- mlebme関数の使い方（今の場合，2PLMを仮定したとする）
  - mlebme(resp = TEST[, 1:5], ip = para.2PL$est, method = "BM")
    - resp: テストデータを指定する引数
    - ip: テストに含まれる各項目の項目母数を指定する引数
      - ”para.2PL$est”と指定することでデータに対し2PLMを仮定したときの項目母数の推定値が各項目の項目母数として指定される
      - 必ず，項目母数を推定した後に，潜在特性値の推定を行う
    - method: どの推定法を用いて潜在特性値を推定するか
      - "ML"と指定することで最尤推定法による潜在特性値の推定を行う

      ⇒　全問正答/誤答の受験者がいる場合，これらの受験者の推定値が求まらない
      - "BM"と指定することで，潜在特性値が標準正規分布に従っていることを仮定しているという情報を加味して推定が行えるようになり（事前情報），全問正答/誤答の受験者に対しても推定値を得ることができる（ベイズ推定）
  - theta.est<-mlebme(resp=TEST[,1:5],ip=para.2PL$est,method="BM")

  ⇒　潜在特性値の推定結果がtheta.estに

R Consoleに”theta.est”と打ち込んでEnter
⇒　潜在特性値の推定結果が表示

```
> theta.est
                  est        sem n
 [1,] -1.89539187 0.7954829 5
 [2,] -1.89539187 0.7954829 5
 [3,] -1.89539187 0.7954829 5
 [4,] -1.47931420 0.7960948 5
 [5,] -1.47931420 0.7960948 5
 [6,] -1.47931420 0.7960948 5
 [7,] -1.47931420 0.7960948 5
 [8,] -1.47931420 0.79
 [9,] -1.47931420 0.79
[10,] -1.45960339 0.79
[11,] -1.45960339 0.79
```

- １列目：推定値
- ２列目：標準誤差
- ３列目：解答した項目数

局所独立性の検討

局所独立性の仮定とその検討方法

- 光永先生の資料より
  - IRTに基づきテストデータや質問紙への回答を分析する

  ⇒　局所独立性が仮定されている
- IRTにおける局所独立性の検討
  - Q_3統計量に基づいて行われることが多い
- Q_3統計量
  - 各項目への回答（観測値）からその期待値を引き，得られた残差得点間の相関を求めることで得られる
  - たとえば今の場合，Item1の残差得点d_1は

$$d_1 = u_1 - \hat{P}_1(\hat{\theta})$$

  - u_1: Item1への解答結果（正答なら1，誤答なら0）
  - $\hat{P}_1(\hat{\theta})$: 項目母数と潜在特性の推定値から計算される正答確率

  ⇒　この残差得点間の相関がQ_3統計量に

RでQ_3統計量を計算する方法

- RでQ_3統計量を計算する
  1. irf関数（正答確率を計算する）を利用して正答確率を推定

  ⇒ irf関数はirtoysパッケージの中に入っている関数

  2. 推定された正答確率とテストデータより，残差得点を算出
  3. cor関数（相関を計算する）を利用してQ_3統計量の値を計算
- irf関数の使い方（今の場合，2PLMを仮定したとする）
  - irf(ip = para.2PL$est, x = theta.est[, 1])
    - ip: テストに含まれる各項目の項目母数を指定する引数
      - ”para.2PL$est”と指定することでデータに対し2PLMを仮定したときの項目母数の推定値が各項目の項目母数として指定される
    - x: 各受験者の潜在特性値を指定する引数
      - ”theta.est[, 1]”と指定することでデータに対し2PLMを仮定したときの潜在特性の推定値が指定される
  - P <- irf(ip = para.2PL$est, x = theta.est[, 1])

  ⇒ 正答確率の推定値がPに

R Consoleに”P”と打ち込んでEnter
⇒　推定結果が表示される

```
$f
          Item1     Item2     Item3     Item4     Item5
 [1,] 0.7700558 0.4061064 0.1917689 0.4949268 0.6915701
 [2,] 0.7700558 0.4061064 0.1917689 0.4949268 0.6915701
 [3,] 0.7700558 0.4061064 0.1917689 0.4949268 0.6915701
 [4,] 0.8252089 0.4801900 0.2557741 0.5661590 0.7466871
 [5,] 0.8252089 0.4801900 0.2557741 0.5661590 0.7466871
 [6,] 0.8252089 0.4801900 0.2557741 0.5661590 0.7466871
 [7,] 0.8252089 0.4801900 0.2557741 0.5661590 0.7466871
 [8,] 0.8252089 0.4801900 0.2557741 0.5661590 0.7466871
 [9,] 0.8252089 0.4801900 0.2557741 0.5661590 0.7466871
[10,] 0.8275                                 94896 0.7491304
[11,] 0.8275                                 94896 0.7491304
```

- $fに正答確率の推定値
  - 行：受験者
  - 列：項目
  - セル：正答確率の推定値

RでQ_3統計量を計算する方法2

- 正答確率が求まったら，今の場合，以下のようにする
  - d <- TEST[, 1:5] - P$f
    - P$f: “P$f”と指定することで，正答確率の推定値が抽出される

  ⇒ テストデータ（TEST[,1:5]）から正答確率を引いた残差得点がdに
  - Q3 <- cor(x = d, y = d, use = "pairwise.complete.obs")

  ⇒ 残差得点間の相関（Q_3統計量の値）が計算されQ3に

  ⇒ 絶対値が0.2以上となっている項目間には局所依存性の疑い（Chen & Thissen, 1997）
    - 局所依存性：局所独立性が満たされていない状態のこと

  ⇒ あくまで一つの基準に過ぎないので注意すること

R Consoleに”Q3”と打ち込んでEnter
⇒ Q_3統計量の値が表示

```
> Q3
              Item1       Item2       Item3        Item4        Item5
Item1  1.00000000 -0.04142824 -0.04101429 -0.064167975 -0.062538809
Item2 -0.04142824  1.00000000 -0.11322248 -0.097060194 -0.029585197
Item3 -0.04101429 -0.11322248  1.00000000 -0.092262203 -0.104216701
Item4 -0.06416797 -0.09706019 -0.09226220  1.000000000 -0.003656669
Item5 -0.06253881 -0.02958520 -0.10421670 -0.003656669  1.000000000
```

- 得られたQ_3統計量の値

⇒ 0.2を超えるものはない

⇒ 局所独立性が成立していることを示唆

項目適合度の検討

Rで項目適合度を検討する方法

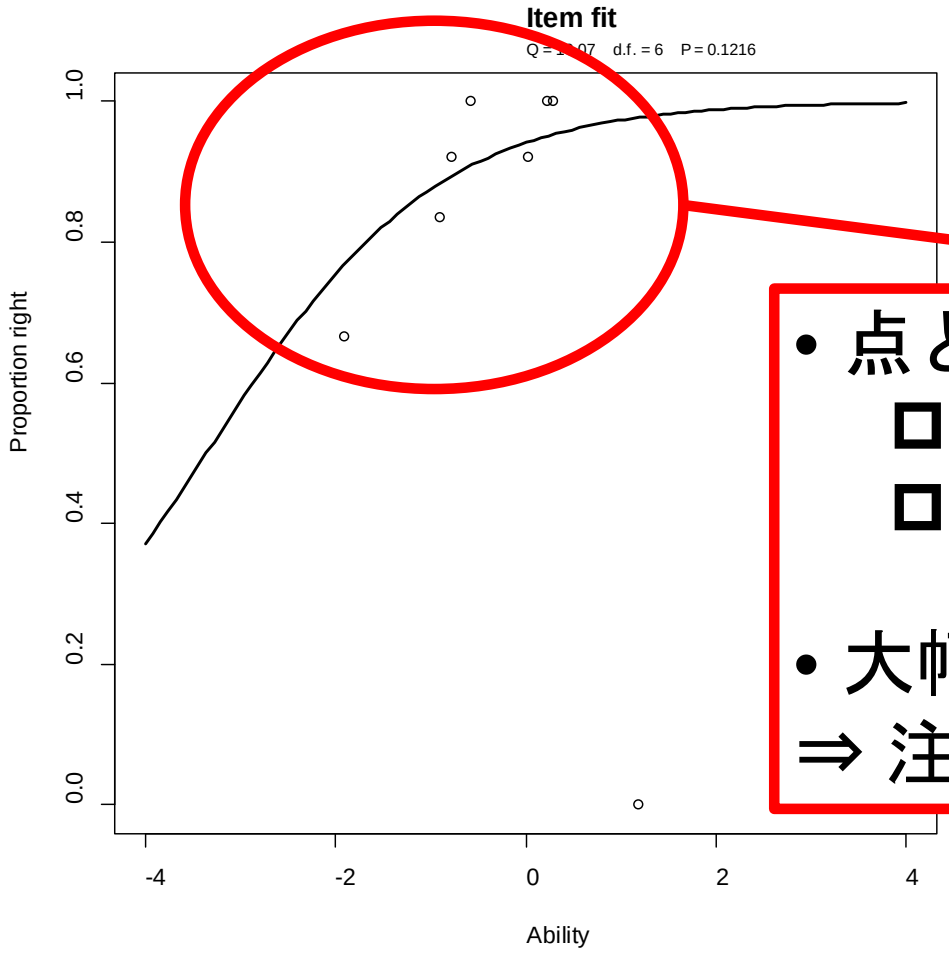
- 光永先生の資料より
  - IRTにおいては項目反応モデルへの適合度を検討することも重要
- Rで項目適合度を検討する
  - itf関数（項目適合度を検討する）を利用して項目適合度を検討

  ⇒　itf関数はirtoysパッケージの中に入っている関数
- itf関数の使い方（今の場合，2PLMを仮定したとする）
  - itf(resp=TEST[, 1:5], item=1, ip=para.2PL$est, theta=theta.est[, 1])
    - resp: テストデータを指定する引数
    - item: どの項目の項目適合度を検討するか指定する引数
      - ”1”と指定することで1番目の項目の項目適合度が検討される
    - ip: テストに含まれる各項目の項目母数を指定する引数
      - ”para.2PL$est”と指定することでデータに対し2PLMを仮定したときの項目母数の推定値が各項目の項目母数として指定される
    - theta: 各受験者の潜在特性値を指定する引数
      - ”theta.est[, 1]”と指定することでデータに対し2PLMを仮定したときの潜在特性の推定値が指定される

  ⇒　項目適合度の検討結果（検定結果と図）が出力される

先のプログラムを実行する
⇒ 項目適合度の検討結果が表示される

```
>  itf(resp = TEST[, 1:5], item = 1, ip = para.2PL$est, theta = theta.est[, 1])
 Statistic          DF     P-value
10.0741811   6.0000000   0.1215627
```



- “P-value”の値が.05以上
⇒ 適合していることを示唆

- 点と実線の乖離度が指標
  - 実線：2PLMに基づく正答確率
  - 点：潜在特性の値に基づき群分けした時の群毎の正答率
- 大幅に乖離している場合
⇒ 注意が必要

項目特性曲線の作成

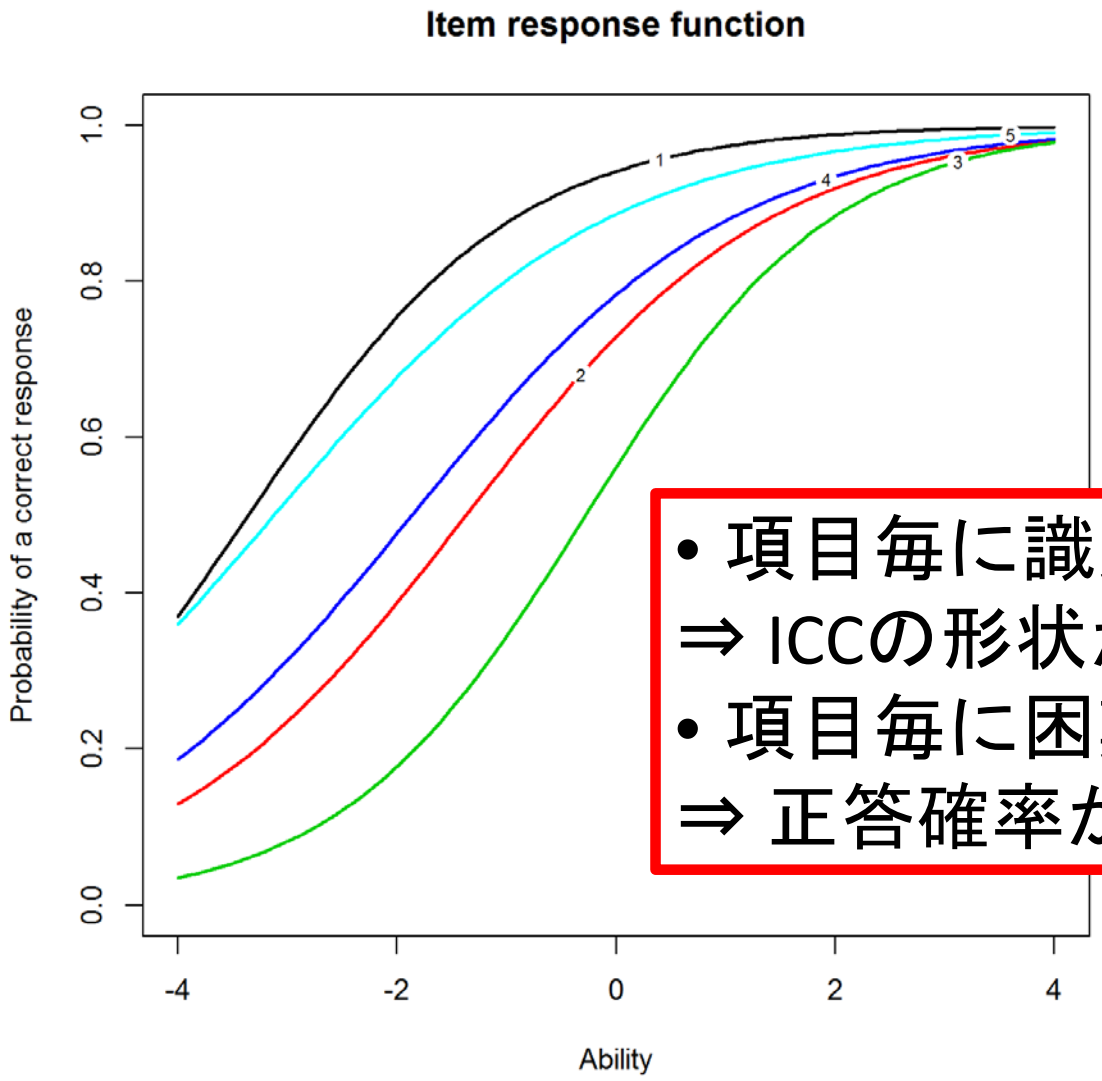
Rで項目特性曲線を作成する方法

- 潜在特性値の値毎に正答確率を計算し，プロットしたもの
  - 項目特性曲線（Item Characteristic Curve，ICC）
- RでICCを作成する
  1. irf関数（正答確率を計算する）を利用して正答確率を計算
  2. plot関数（図を作成する）でICCを作成
- irf関数とplot関数の使い方（今の場合，2PLMを仮定）
  - P <- irf(ip = para.2PL$est)
    - ip: テストに含まれる各項目の項目母数を指定する引数
      - ”para.2PL$est”と指定することでデータに対し2PLMを仮定したときの項目母数の推定値が各項目の項目母数として指定される

  ⇒ 様々な潜在特性値における正答確率がPに
  - plot(x = P, co = NA, label = TRUE)
    - x: irf関数で推定した結果を指定する引数
    - co: ICCの色を指定する引数
      - “NA”とすることで，項目毎に異なる色でICCが描かれる
    - label: ICCに項目の番号を付けるか否か
      - “TRUE（はい）”とすることで各ICCに項目の番号がつく

  ⇒ 各項目のICCが作成される

先のプログラムを実行する
⇒　以下の図が得られる



- 項目毎に識別力母数が異なる
⇒ ICCの形状が異なる
- 項目毎に困難度母数が異なる
⇒ 正答確率が0.5になる点が異なる

テスト特性曲線の作成

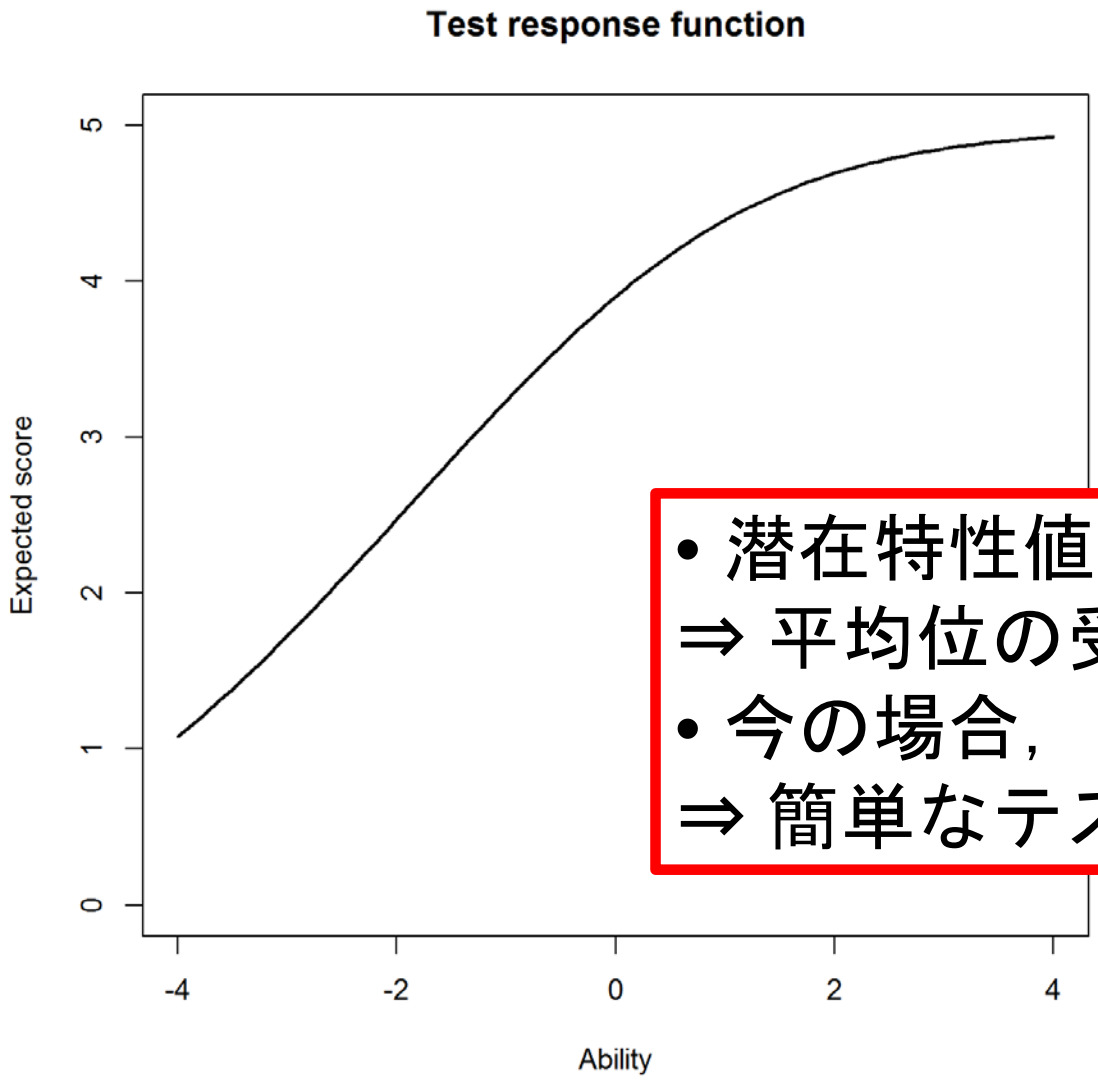
Rでテスト特性曲線を作成する方法

- 潜在特性値の値毎に素点の期待値を計算しプロットする
  - テスト特性曲線（Test Characteristic Curve，TCC）
- RでTCCを作成する
  1. trf関数（素点の期待値を計算する）を利用して期待値を計算
  ⇒ trf関数はirtoysパッケージの中に入っている関数
  2. plot関数（図を作成する）でTCCを作成
- trf関数とplot関数の使い方（今の場合，2PLMを仮定）
  - E <- trf(ip = para.2PL$est)
    - ip: テストに含まれる各項目の項目母数を指定する引数
      - ”para.2PL$est”と指定することでデータに対し2PLMを仮定したときの項目母数の推定値が各項目の項目母数として指定される

  ⇒ 様々な潜在特性値における素点の期待値（Expectation）がEに
  - plot(x = E)
    - x: trf関数で推定した結果を指定する引数

  ⇒ TCCが作成される

先のプログラムを実行する
⇒　以下の図が得られる



- 潜在特性値の平均は０
⇒ 平均位の受験者の期待値は４
- 今の場合，満点は５
⇒ 簡単なテストであることが示唆

TCCの作成を行った論文例

教育心理学研究, 2015, 63, 111—120

大学生を対象とした精神的健康度調査の共通尺度化による比較検討

酒　井　　渉*　　野　口　裕　之**

近年，学生相談において，精度が高く簡便な精神的健康度のスクリーニング検査への要請が高まっている。従来，UPI，GHQ，K 10 の三者は，同じものを測っているのか，互いのカットオフポイントが共通なのかが不明確なまま，ともに「精神的健康度のスクリーニング検査」として用いられてきた。本研究では，この二点を明らかにすることを目的とした。大学新入生 548 名（男子 270 名, 女子 278 名）を受検者とし，UPI，GHQ-30，K 10 を同時施行した。その回答を，相関係数，因子分析，項目反応理論による共通尺度化を用いて分析した。その結果，三者のカットオフポイントは共通尺度上でほぼ同値であった。また三者の特徴が明らかとなった。UPI は日常的な困りごとのレベルで，GHQ-30 はカットオフポイント付近での測定精度が高い。K 10 は少ない項目数で病気を切り分けることができる。三者は特徴に応じた使い分けが可能であるとわかった。

キーワード：項目反応理論，精神的健康度調査，UPI，GHQ，K 10

酒井・野口 (2015)におけるTCCの作成

- 精神的健康度を測定する目的で作成された尺度
  - UPI（30点がカットオフポイント）
  - GHQ（青年期の回答者は13点がカットオフポイント）
  - K10（15点がカットオフポイント）
- 上記の3つの尺度を大学の新入生599名に実施
  - 欠測値の無い548名分のデータを分析
- K10が5件法の質問紙
  - 項目反応モデルとしてGRMを適用して項目母数を推定

  ⇒ 得られた項目母数の推定値よりTCCを作成

  ⇒ 尺度間の統計的性質の比較が容易に（次ページ）

酒井・野口 (2015)におけるTCC

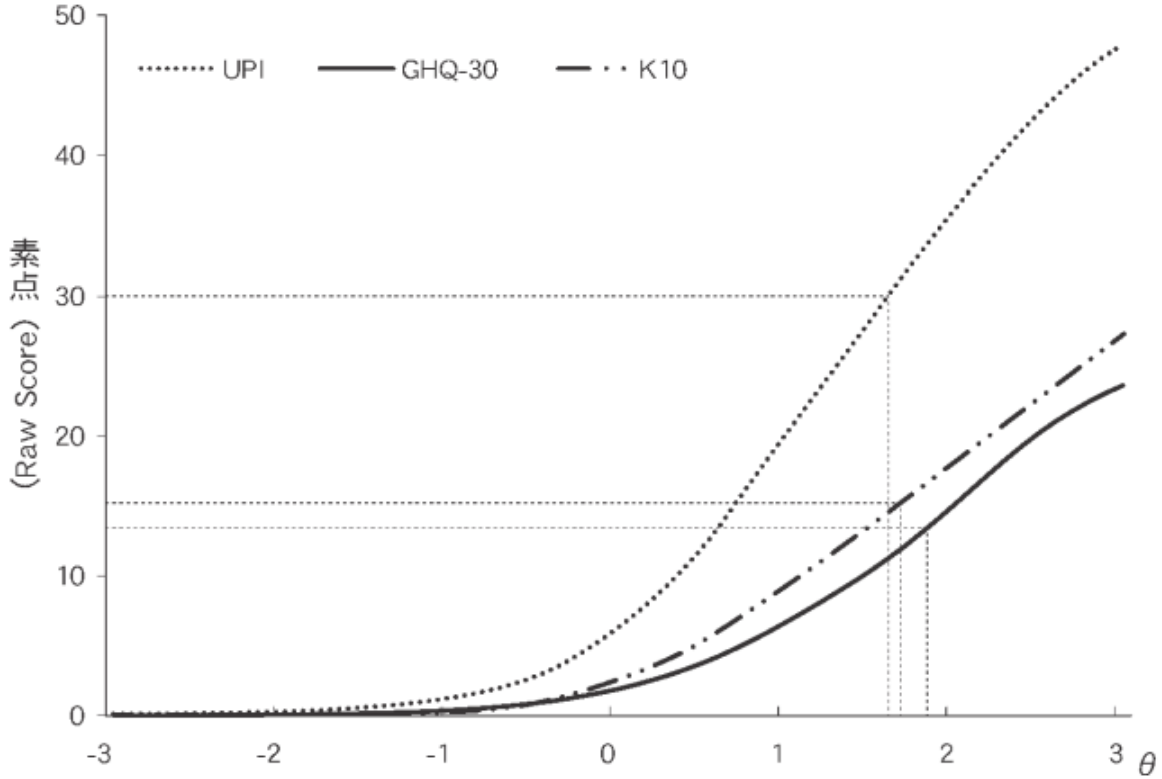


Figure 1 テスト特性曲線（Test Characteristic Curve）

- 各尺度のカットオフポイント

⇒ 大体同じ潜在特性値に対応

⇒ どの尺度でスクリーニングしても大体同じ結果に

テスト情報曲線の作成

Rでテスト情報曲線を作成する方法

- 潜在特性値の値毎にテスト情報量を計算しプロットする
  - テスト情報曲線（Test Information Curve，TIC）
- RでTICを作成する
  1. tif関数（テスト情報量を計算する）を利用しテスト情報量を計算

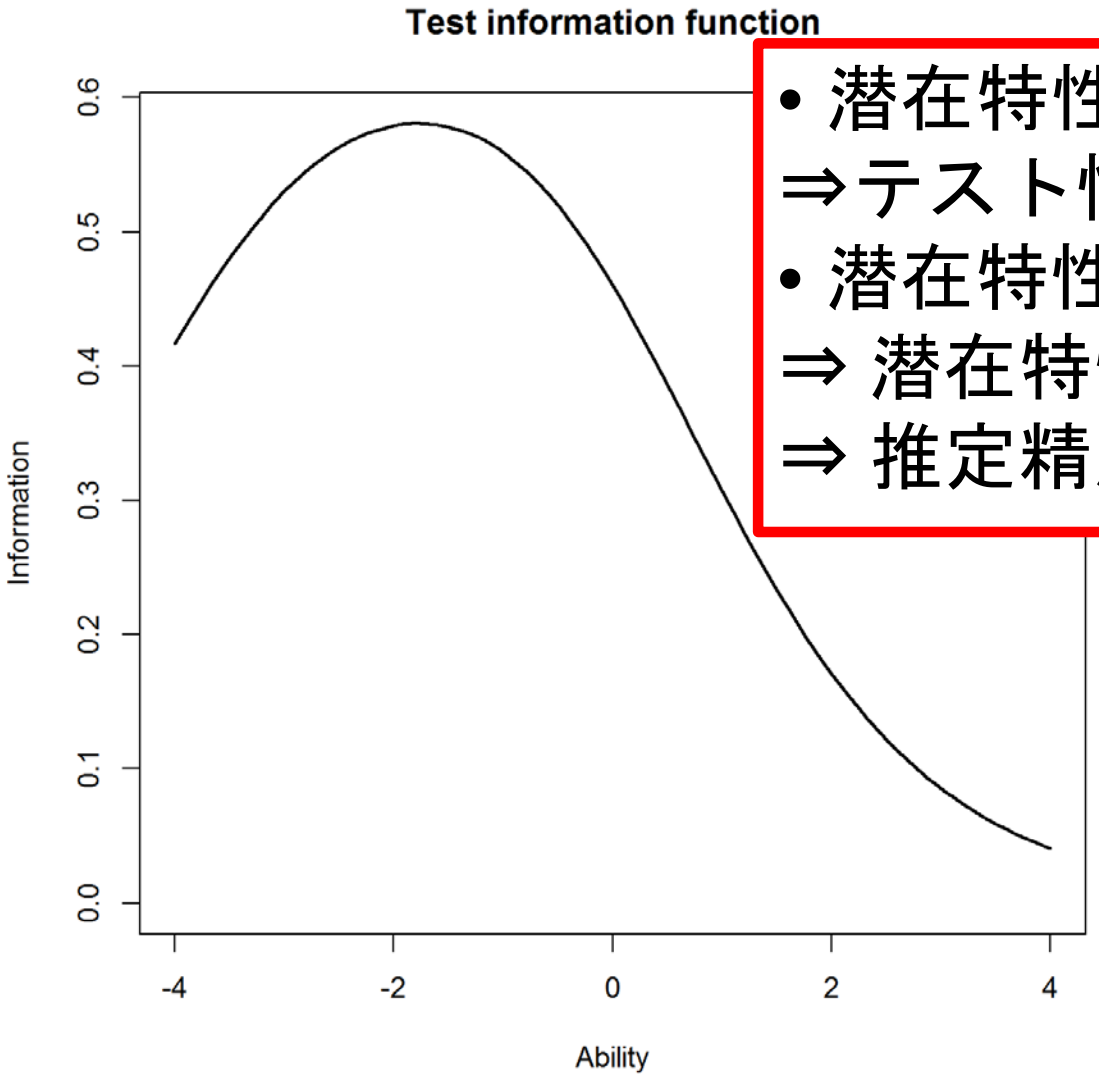
  ⇒ tif関数はirtoysパッケージの中に入っている関数

  2. plot関数（図を作成する）でTICを作成
- tif関数とplot関数の使い方（今の場合，2PLMを仮定）
  - I <- tif(ip = para.2PL$est)
    - ip: テストに含まれる各項目の項目母数を指定する引数
      - ”para.2PL$est”と指定することでデータに対し2PLMを仮定したときの項目母数の推定値が各項目の項目母数として指定される

  ⇒ 様々な潜在特性値におけるテスト情報量（Information）がIに
  - plot(x = I)
    - x: tif関数で推定した結果を指定する引数

  ⇒ TICが作成される

先のプログラムを実行する
⇒　以下の図が得られる



- 潜在特性値が-2の辺り

⇒テスト情報量が最も大きくなる

- 潜在特性値は平均０，分散１

⇒ 潜在特性値が低い受験者

⇒ 推定精度が最も高くなる

TICの作成を行った論文例

教育心理学研究, 2015, 63, 111—120

大学生を対象とした精神的健康度調査の共通尺度化による比較検討

酒　井　　渉*　　野　口　裕　之**

近年，学生相談において，精度が高く簡便な精神的健康度のスクリーニング検査への要請が高まっている。従来，UPI，GHQ，K 10 の三者は，同じものを測っているのか，互いのカットオフポイントが共通なのかが不明確なまま，ともに「精神的健康度のスクリーニング検査」として用いられてきた。本研究では，この二点を明らかにすることを目的とした。大学新入生 548 名（男子 270 名，女子 278 名）を受検者とし，UPI，GHQ-30，K 10 を同時施行した。その回答を，相関係数，因子分析，項目反応理論による共通尺度化を用いて分析した。その結果，三者のカットオフポイントは共通尺度上でほぼ同値であった。また三者の特徴が明らかとなった。UPI は日常的な困りごとのレベルで，GHQ-30 はカットオフポイント付近での測定精度が高い。K 10 は少ない項目数で病気を切り分けることができる。三者は特徴に応じた使い分けが可能であるとわかった。

キーワード：項目反応理論，精神的健康度調査，UPI，GHQ，K 10

酒井・野口 (2015)におけるTIC

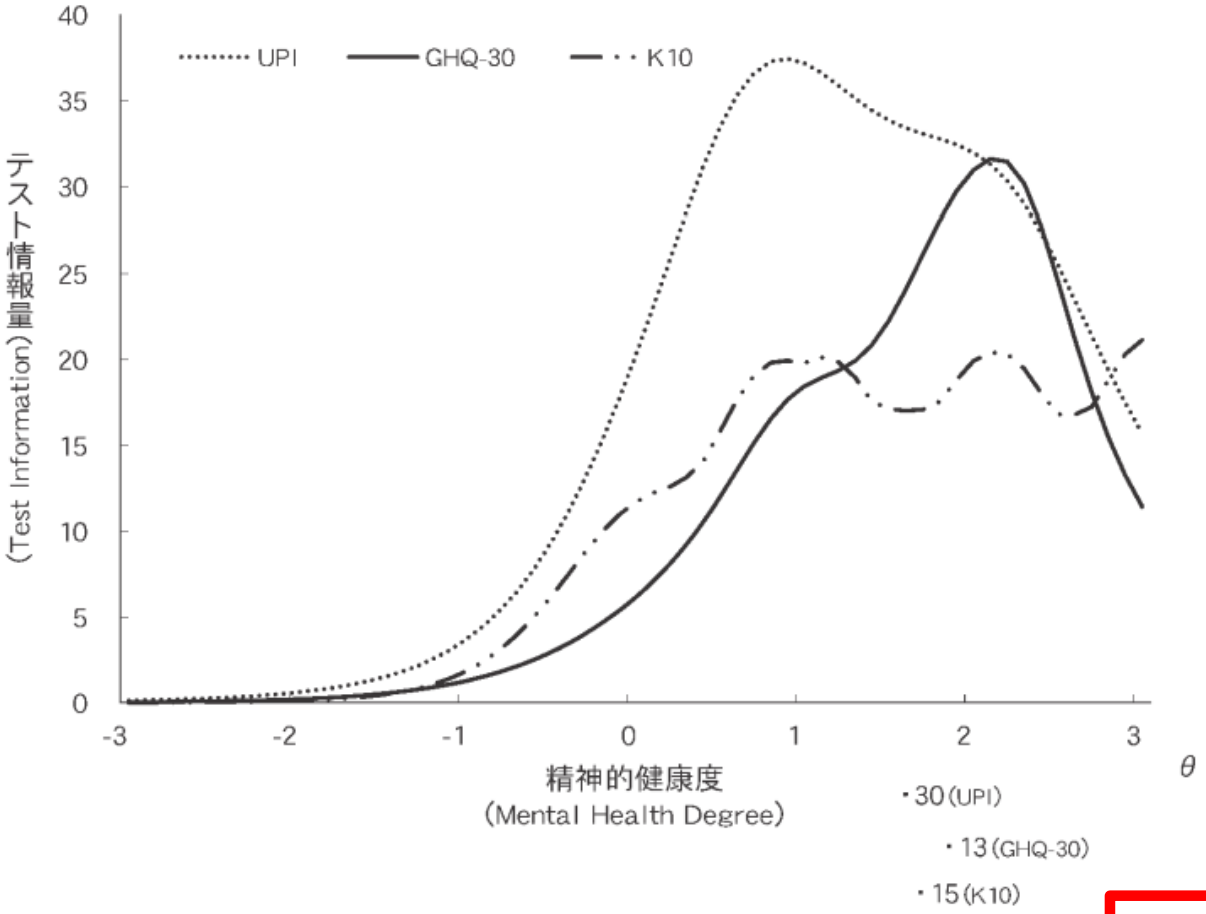


Figure 2 テスト情報曲線（Test Information Curve）

注：下部の「・30（UPI）」等はカットオフポイントを示す。
（「・30(UPI)」etc. is cut-off point at lower part.）

- 異なる尺度間の比較
⇒IRTにより容易に

酒井・野口 (2015)におけるTICの作成

- 各尺度におけるテスト情報量
  - UPI: 潜在特性値が0.9の付近で最大
  - ⇒　日常的な困りごとのレベルでの測定精度が高い
  - GHQ: 潜在特性値が2.2の付近で最大
  - ⇒　面接調査が必要となるレベルでの測定精度が高い
  - K10: 潜在特性値が0.9付近以上で一定の値
  - ⇒　日常的な困りごとのレベル以上での測定精度が高い
- カットオフポイント付近におけるテスト情報量
  - UPI（56項目）: 30程度
  - GHQ（30項目）: 25程度
  - K10（10項目）: 16以上
  - ⇒　K10の測定効率が最も高い
- テスト情報量を利用した短縮版の作成
  - テスト情報量の減少を抑えつつ項目を取捨選択
  - ⇒　測定効率の最大化

質問紙データの解析例

質問紙データの例

- psychパッケージの中に入っている実データを使用
  - SAPAに含まれる自己報告式の性格特性項目への回答
    - 回答者数：2800人
    - 項目数：外向性に関する5項目
    - 6件法のデータ
      - １：とても当てはまらない（Very Inaccurate）
      - …
      - ６：とてもあてはまる（Very Accurate）

  ⇒　このデータに対してGRMを適用し，IRTに基づく評価を行う

- SAPA
  - The Synthetic Aperture Personality Assessment (SAPA)と呼ばれるウェブベースの性格特性調査プロジェクト
  - 今回使用しなかった変数も含め，データについてのより詳しい解説についてはR Consoleに”?bfi”と打ち込むことで参照可能
- psychパッケージをロードした後
  - R Consoleに”data(bfi)”と入力しても入手可能

質問紙データの例

ID	E1	E2	E3	E4	E5
61617	4	4	3	4	4
61618	6	6	6	4	3
61620	5	3	4	4	5
61621	2	4	4	4	4
61622	5	5	5	4	5
61623	5	6	6	5	6
61624	3	4	4	5	5
61629	4	1	4	2	1
61630	2	4	NA	4	3
61633	5	5	4	5	5
61634	6	4	2	5	4

Rに質問紙データを読み込む方法

- Rに質問紙データ（CSVファイル）を読み込む
  - read.csv関数（CSVファイルを読み込む）を利用して読み込み
- read.csv関数の使い方（今の場合）
  - read.csv(file = "bfi.csv", header = TRUE, row.names = 1)
    - File: 質問紙データの入っているCSVファイルの名前を指定する引数
      - 今の場合，利用するデータは"bfi.csv"に入っているので，これを指定する
    - header: CSVファイルの１行目に変数名が記載されているか否か
      - ”TRUE（はい）”と指定することでCSVファイルの１行目が変数名と認識される
    - row.names: 回答者IDが何列目に記載されているかを指定する引数
      - ”1”と指定することでCSVファイルの１列目が回答者IDと認識される
  - SCALE <- read.csv(file = "bfi.csv", header = TRUE, row.names = 1)

⇒　読み込んだCSVファイルの質問紙（SCALE）データがSCALEに

R Consoleに“SCALE”と打ち込んでEnter
⇒　読み込んだデータが表示される

```
> SCALE
      E1 E2 E3 E4 E5
61617  4  4  3  4  4
61618  6  6  6  4  3
61620  5  3  4  4  5
61621  2  4  4  4  4
61622  5  5  5  4  5
61623  5  6  6  5  6
61624  3  4  4  5  5
61629  4  1  4  2  1
61630  2  4 NA  4  3
61633  5  5  4  5  5
61634  6  4  2  5  4
```

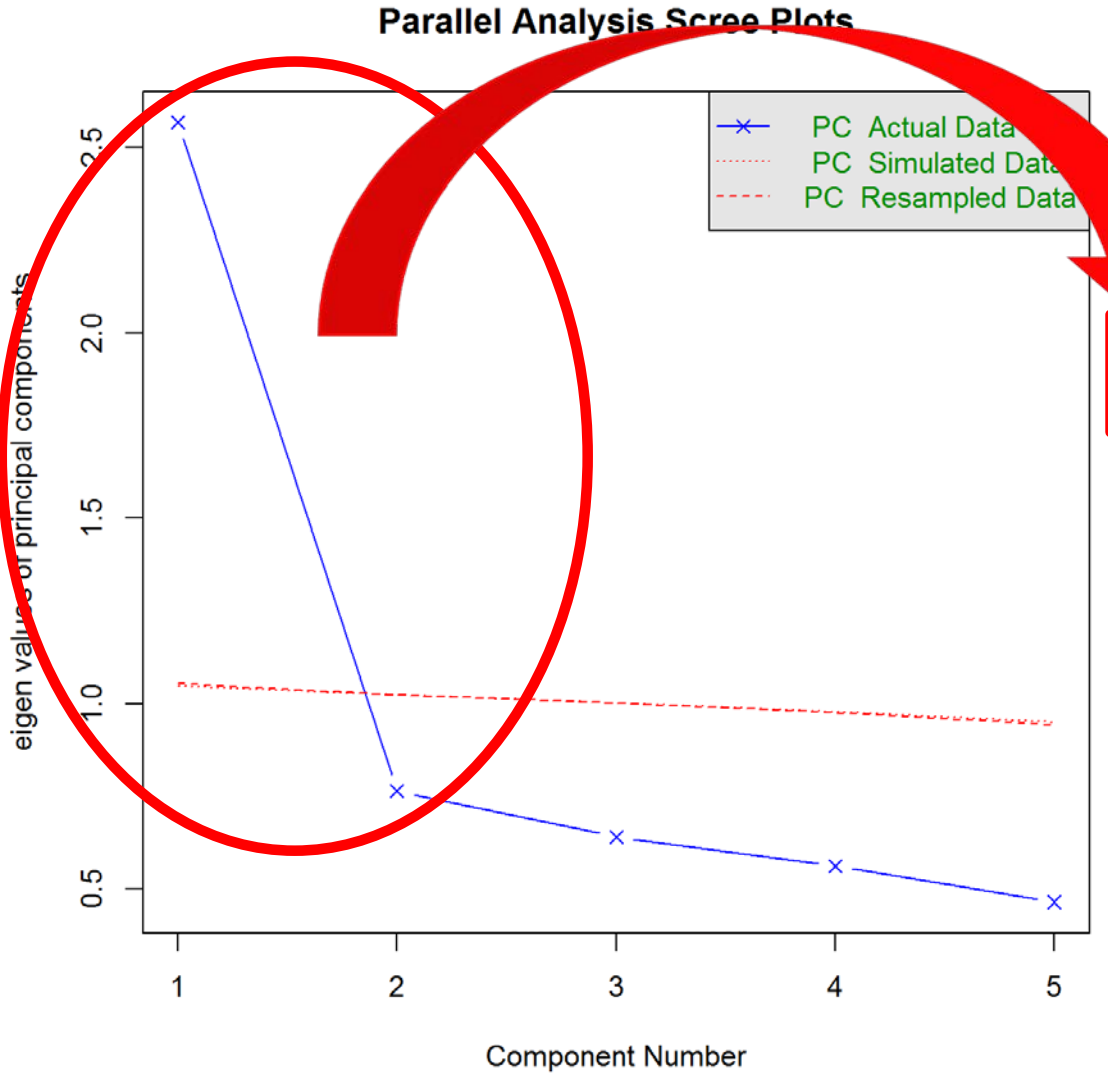
１次元性の検討

Rでのスクリープロットの作成と平行分析の実施方法

- Rでのスクリープロットの作成と平行分析の実施
  - fa.parallel関数（平行分析を実施する）でまとめて実行可能

  ⇒　fa.parallel関数はpsychパッケージの中に入っている関数
- fa.parallel関数の使い方（今の場合）
  - fa.parallel(x = SCALE, fa = "pc")
    - x: 質問紙データを指定する引数
    - fa: 固有値分解の対象となる行列を指定する引数
      - 項目得点間の相関行列を固有値分解するので"pc"と指定

先のプログラムを実行する
⇒　平行分析の結果等が表示される



１次元性が示唆された

段階反応モデルの項目母数の推定

RでGRMの項目母数を求める方法

- 光永先生の資料より
  - IRTにおいては項目母数の値よりその項目の特性を把握
- RでGRMを仮定したときの項目母数を周辺最尤推定する
  - grm関数（GRMの項目母数を推定する）を利用し項目母数を推定

  ⇒　grm関数はltmパッケージの中に入っている関数
- grm関数の使い方（今の場合）
  - grm(data = SCALE)
    - data: 質問紙データを指定する引数
      - 質問紙データは行が受験者，列が項目を表すようにする
  - para.grm <- grm(data = SCALE)

  ⇒　項目母数の推定結果がpara.grmに

R Consoleに”para.grm”と打ち込んでEnter
⇒　推定結果が表示される

```
> para.grm

Call:
grm(data = SCALE)

Coefficients:
    Extrmt1  Extrmt2  Extrmt3  Extrmt4  Extrmt5  Dscrmn
E1   -2.099   -1.163   -0.456    0.087    1.059   1.494
E2   -1.712   -0.934   -0.185    0.201    1.100   2.176
E3   -2.628   -1.569   -0.770    0.426    1.822   1.386
E4   -2.227   -1.397   -0.914   -0.321    0.842   1.981
E5   -3.367   -2.128   -1.339   -0.243    1.339   1.175
```

- Coefficientsに項目母数の推定値
  - １列目：２以上と反応する確率に関する困難度母数
  - …
  - ５列目：６と反応する確率に関する困難度母数
  - ６列目：1.7×識別力母数

GRMの項目母数推定を行った論文例

教育心理学研究, 2015, 63, 111—120

大学生を対象とした精神的健康度調査の共通尺度化による比較検討

酒　井　　渉*　　野　口　裕　之**

近年，学生相談において，精度が高く簡便な精神的健康度のスクリーニング検査への要請が高まっている。従来，UPI，GHQ，K 10 の三者は，同じものを測っているのか，互いのカットオフポイントが共通なのかが不明確なまま，ともに「精神的健康度のスクリーニング検査」として用いられてきた。本研究では，この二点を明らかにすることを目的とした。大学新入生 548 名（男子 270 名, 女子 278 名）を受検者とし，UPI，GHQ-30，K 10 を同時施行した。その回答を，相関係数，因子分析，項目反応理論による共通尺度化を用いて分析した。その結果，三者のカットオフポイントは共通尺度上でほぼ同値であった。また三者の特徴が明らかとなった。UPI は日常的な困りごとのレベルで，GHQ-30 はカットオフポイント付近での測定精度が高い。K 10 は少ない項目数で病気を切り分けることができる。三者は特徴に応じた使い分けが可能であるとわかった。

キーワード：項目反応理論，精神的健康度調査，UPI，GHQ，K 10

酒井・野口 (2015)における項目母数の推定

- 精神的健康度を測定する目的で作成された尺度
  - UPI（30点がカットオフポイント）
  - GHQ（青年期の受験者は13点がカットオフポイント）
  - K10（15点がカットオフポイント）
- 上記の3つの尺度を大学の新入生599名に実施
  - 欠測値の無い548名分のデータを分析
- K10が5件法の質問紙
  - 項目反応モデルとしてGRMを適用して項目母数を推定

酒井・野口 (2015)におけるGRMの項目母数の推定値

Table 4　項目パラメータ一覧（K10）

$N = 548$

項目番号	識別力	位置			
		1	2	3	4
1	1.89 (0.19)	−0.16 (0.08)	0.68 (0.08)	2.13 (0.18)	3.31 (0.39)
2	1.61 (0.19)	0.46 (0.09)	1.46 (0.14)	2.68 (0.27)	3.59 (0.47)
3	1.99 (0.31)	1.25 (0.12)	2.04 (0.19)	3.02 (0.36)	3.71 (0.62)
4	1.82 (0.26)	1.05 (0.11)	1.82 (0.18)	2.96 (0.34)	4.29 (0.95)
5	1.50 (0.16)	−0.16 (0.09)	1.07 (0.12)	2.75 (0.28)	3.64 (0.46)
6	1.47 (0.21)	1.27 (0.15)	2.47 (0.28)	3.85 (0.55)	4.98 (1.09)
7	2.24 (0.20)	−0.02 (0.07)	0.86 (0.08)	2.04 (0.16)	3.03 (0.33)
8	2.22 (0.23)	0.61 (0.07)	1.34 (0.11)	2.23 (0.19)	3.16 (0.37)
9	1.47 (0.19)	0.84 (0.11)	1.93 (0.21)	3.06 (0.38)	4.50 (0.90)
10	1.44 (0.22)	1.18 (0.14)	2.35 (0.27)	3.24 (0.42)	4.05 (0.67)
平均	1.77		2.22		
標準偏差	0.30		1.33		

注：（　）内は標準誤差を示す。

項目反応カテゴリ特性曲線の作成

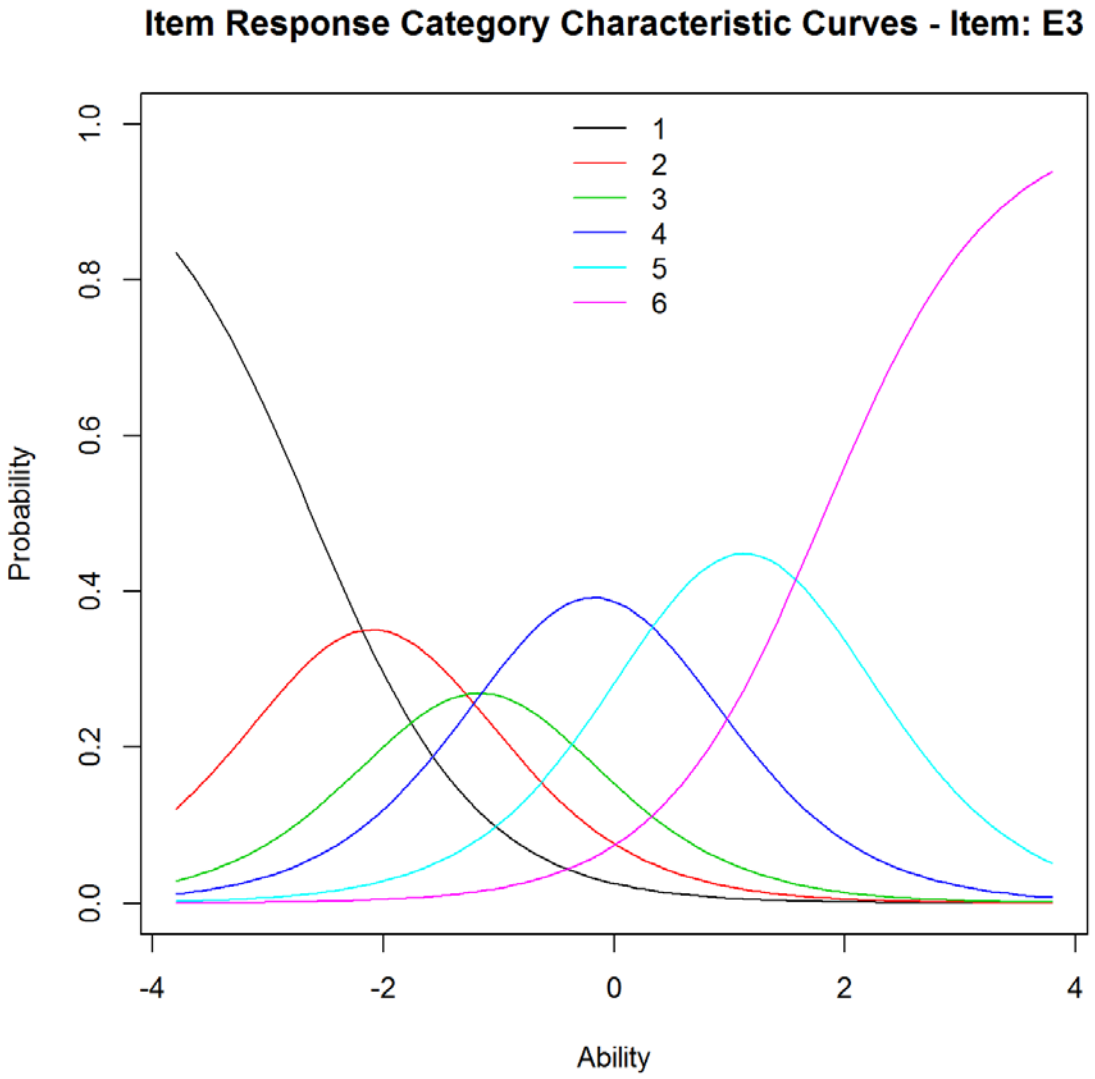
Rで項目反応カテゴリ特性曲線を作成する方法

- Rで項目反応カテゴリ特性曲線を作成する
  - plot関数（図を作る）を利用して項目反応カテゴリ特性曲線を作成
- plot関数の使い方（今の場合）
  - plot(x = para.grm, type = "ICC", legend = TRUE)
  - x: grm関数での推定結果を指定する引数
    - type: GRMに関するどのような曲線を描くのか指定する引数
      - "ICC"と指定することで，項目反応カテゴリ特性曲線を作成する
    - legend: 作成した項目反応カテゴリ特性曲線に凡例をつけるか否か
      - “TRUE（はい）”とすることで各曲線がどのカテゴリに対応しているかの凡例がつくことになる
- 先のコードを実行する
  - 各項目の項目反応カテゴリ特性曲線が作成される

⇒　1項目ずつ結果が表示されるのでEnterを押して切り替える

先のプログラムを実行する

⇒　たとえば，以下の図が得られる



等化の実行例

本セクションで用いるデータ

- plinkパッケージの中に入っているデータを使用
  - 異なる2つのテスト（oldとnew）を異なる2つの集団が受験
    - テストに含まれる項目数：それぞれ36項目

  ⇒ このデータに対して3PLMを当てはめて項目母数を推定
    - a.old: テストoldにおける識別力母数の推定値
    - b.old: テストoldにおける困難度母数の推定値
    - c.old: テストoldにおける当て推量母数の推定値
    - a.new: テストnewにおける識別力母数の推定値
    - b.new: テストnewにおける困難度母数の推定値
    - c.new: テストnewにおける当て推量母数の推定値
  - ２つのテストに含まれる共通項目
    - 3，6，9，12，15，18，21，24，27，30，33，36番目の項目（計12個）

  ⇒ newの項目母数をoldのものと比較可能にするため，等化する
- plinkパッケージを読み込んだ後
  - R Consoleに”KB04”と入力しても入手可能

Item	a.old	b.old	c.old	a.new	b.new	c.new
Item1	0.8704	−1.4507	0.1576	0.5496	−1.796	0.1751
Item2	0.4628	−0.407	0.1094	0.7891	−0.4796	0.1165
Item3	0.4416	−1.3349	0.1559	0.4551	−0.7101	0.2087
Item4	0.5448	−0.9017	0.1381	1.4443	0.4833	0.2826
Item5	0.62	−1.4865	0.2114	0.974	−0.168	0.2625
Item6	0.573	−1.321	0.1913	0.5839	−0.8567	0.2038
Item7	1.1752	0.0691	0.2947	0.8604	0.4546	0.3224
Item8	0.445	0.2324	0.2723	1.1445	−0.1301	0.2209
Item9	0.5987	−0.7098	0.1177	0.7544	0.0212	0.16
Item10	0.8479	−0.4253	0.1445	0.917	1.0139	0.3648
Item11	1.032	−0.8184	0.0936	0.9592	0.7218	0.2399
Item12	0.6041	−0.3539	0.0818	0.6633	0.0506	0.124
Item13	0.8297	−0.0191	0.1283	1.2324	0.4167	0.2535
Item14	0.7252	−0.3155	0.0854	1.0492	0.7882	0.1569
Item15	0.9902	0.532	0.3024	1.069	0.961	0.2986
Item16	0.7749	0.5394	0.2179	0.9193	0.6099	0.2521
Item17	0.5942	0.8987	0.2299	0.8935	0.5128	0.2273
Item18	0.8081	−0.1156	0.0648	0.9672	0.195	0.0535
Item19	0.964	−0.1948	0.1633	0.6562	0.3953	0.1201
Item20	0.7836	0.3506	0.1299	1.0556	0.9481	0.2036
Item21	0.414	2.5538	0.241	0.3479	2.2768	0.1489
Item22	0.7618	−0.1581	0.1137	0.8432	1.0601	0.2332
Item23	1.1959	0.5056	0.2397	1.1142	0.5826	0.0644
Item24	1.3554	0.5811	0.2243	1.4579	1.0241	0.2453
Item25	1.1869	0.6229	0.2577	0.5137	1.379	0.1427
Item26	1.0296	0.3898	0.1856	0.9194	1.0782	0.0879
Item27	1.0417	0.9392	0.1651	1.8811	1.4062	0.1992
Item28	1.2055	1.135	0.2323	1.5045	1.5093	0.1642
Item29	0.9697	0.6976	0.107	0.9664	1.5443	0.1431
Item30	0.6336	1.896	0.0794	0.702	2.2401	0.0853
Item31	1.0822	1.3864	0.1855	1.2651	1.8759	0.2443
Item32	1.0195	0.9197	0.1027	0.8567	1.714	0.0865
Item33	1.1347	1.079	0.063	1.408	1.5556	0.0789
Item34	1.1948	1.8411	0.0999	0.5808	3.4728	0.1399
Item35	1.1961	2.0297	0.0832	0.9257	3.1202	0.109
Item36	0.9255	2.1337	0.1259	1.2993	2.1589	0.1075

- 黄色の部分
⇒ 共通項目
- 白色の部分
⇒ oldとnewで別の項目

Rに項目母数の推定値を読み込む方法

- Rに項目母数の推定値（CSVファイル）を読み込む
  - read.csv関数（CSVファイルを読み込む）を利用して読み込み
- read.csv関数の使い方（今の場合）
  - read.csv(file = "PARA.csv", header = TRUE, row.names = 1)
    - file: 項目母数の推定値が入っているCSVファイルの名前を指定する引数
      - 今の場合，推定値は"PARA.csv"に入っているので，これを指定する
    - header: CSVファイルの１行目に変数名が記載されているか否か
      - ”TRUE（はい）”と指定することでCSVファイルの１行目が変数名と認識される
    - row.names: 項目番号が何列目に記載されているかを指定する引数
      - ”1”と指定することでCSVファイルの１列目が項目番号と認識される
  - PARA <- read.csv(file = "PARA.csv", header = TRUE, row.names = 1)

⇒　読み込んだ項目母数（PARAmeter）の推定値がPARAに

R Consoleに“PARA”と打ち込んでEnter
⇒　読み込んだデータが表示される

```
> PARA
        a.old   b.old  c.old  a.new   b.new  c.new
Item1  0.8704 -1.4507 0.1576 0.5496 -1.7960 0.1751
Item2  0.4628 -0.4070 0.1094 0.7891 -0.4796 0.1165
Item3  0.4416 -1.3349 0.1559 0.4551 -0.7101 0.2087
Item4  0.5448 -0.9017 0.1381 1.4443  0.4833 0.2826
Item5  0.6200 -1.4865 0.2114 0.9740 -0.1680 0.2625
Item6  0.5730 -1.3210 0.1913 0.5839 -0.8567 0.2038
Item7  1.1752  0.0691 0.2947 0.8604  0.4546 0.3224
Item8  0.4450  0.2324 0.2723 1.1445 -0.1301 0.2209
Item9  0.5987 -0.7098 0.1177 0.7544  0.0212 0.1600
Item10 0.8479 -0.4253 0.1445 0.9170  1.0139 0.3648
Item11 1.0320 -0.8184 0.0936 0.9592  0.7218 0.2399
Item12 0.6041 -0.3539 0.0818 0.6633  0.0506 0.1240
Item13 0.8297 -0.0191 0.1283 1.2324  0.4167 0.2535
```

Rで等化を実行するための方法

- 先のようなデータがあるときに，Rで等化を実行する
  - “plink”パッケージのplink関数等を使用すればよい

  ⇒　具体的には，以下の手続きを実行

  1. plink関数に引き渡すデータを作成する
  2. plink関数を利用し等化係数を計算
  3. plink関数で求めた等化係数を利用し，等化を実行
- Rでの等化の手続き
  - 今までの分析手続きに比べると手続きが複雑

  ⇒　赤枠で囲まれている部分をそのまま実行すれば実行可能
- plinkパッケージを読み込んだ後に先ほどのgrm関数を使用
  - grm関数による分析ができなくなる

  ⇒　分析の実行順序に注意が必要

plink関数に引き渡すデータの作成1

1. まずは，plink関数に引き渡す項目母数を用意する
   - para.old <- PARA[, 1:3]

   ⇒ テストoldの項目母数（推定値，PARA[, 1:3]）がpara.oldに
   - para.new <- PARA[, 4:6]

   ⇒ テストnewの項目母数（推定値，PARA[, 4:6]）がpara.newに
   - para.on <- list(para.old, para.new)

   ⇒ para.oldとpara.newを1つのリスト（list）にまとめる

R Consoleに“para.on”と打ち込んでEnter
⇒　para.oldとpara.newがpara.onの中に

```
>  para.on↓
[[1]]↓
          a.old    b.old  c.old
Item1   0.8704 -1.4507 0.1576
Item2   0.4628 -0.4070 0.1094
Item3   0.4416 -1.3349 0.1559
Item4   0.5448 -0.9017 0.1381
Item5   0.6200 -1.4865 0.2114
Item6   0.5730 -1.3210 0.1913
～中略～↓
↓
[[2]]↓
          a.new    b.new  c.new
Item1   0.5496 -1.7960 0.1751
Item2   0.7891 -0.4796 0.1165
Item3   0.4551 -0.7101 0.2087
Item4   1.4443  0.4833 0.2826
Item5   0.9740 -0.1680 0.2625
Item6   0.5839 -0.8567 0.2038
```

plink関数に引き渡すデータの作成2

2. 次に，テストに含まれる各項目のカテゴリ数を指定する
   - その際，rep関数（繰り返しを行う）を利用する

- rep関数の使い方（今の場合）
  - rep(x = 2, times = 36)
    - x: 繰り返す対象を指定する引数
      - 今の場合，2つのテストに含まれる各項目のカテゴリ数（2）を項目数（36）分繰り返したいので，2と指定する
    - times: xを何回繰り返すか指定する引数
      - 今の場合，2つのテストに含まれる各項目のカテゴリ数（2）を項目数（36）分繰り返したいので，36と指定する
  - cat.old <- rep(x = 2, times = 36)

⇒ テストoldの各項目のカテゴリ（category）数がcat.oldに

- cat.new <- rep(x = 2, times = 36)

⇒ テストnewの各項目のカテゴリ（category）数がcat.newに

- cat.on <- list(cat.old, cat.new)

⇒ cat.oldとcat.newを1つのリスト（list）にまとめる

plink関数に引き渡すデータの作成3

3. 続いて，利用した項目反応モデルについて指定する
   - その際，as.poly.mod関数を利用する

- as.poly.mod関数の使い方（今の場合）
  - as.poly.mod(n = 36)
    - n: テストに含まれる項目数を指定する引数
      - 今の場合，両テストとも項目数は36なので，36と指定する

  ⇒ 36項目に3PLM（二値型のモデル）を当てはめたと記録される

  - pm.old <- as.poly.mod(n = 36)

  ⇒ テストoldの全ての項目に二値型のモデルを当てはめたと記録

  - pm.new <- as.poly.mod(n = 36)

  ⇒ テストnewの全ての項目に二値型のモデルを当てはめたと記録

  - pm.on <- list(pm.old, pm.new)

  ⇒ pm.oldとpm.newを1つのリスト（list）にまとめる

plink関数に引き渡すデータの作成4

4. 更に，共通項目の項目番号を指定する
   - その際，cbind関数（列方向に束ねる）を利用する

- cbind関数の使い方（今の場合）
  - common.old <- c(3, 6, 9, 12, 15, 18, 21, 24, 27, 30, 33, 36)

  ⇒ テストoldにおける共通（common）項目の番号がcommon.oldに
  - common.new <- c(3, 6, 9, 12, 15, 18, 21, 24, 27, 30, 33, 36)

  ⇒ テストnewにおける共通項目の番号がcommon.newに
  - common.on <- cbind(common.old, common.new)

  ⇒ common.oldとcommon.newが列方向に束ねられcommon.onに

R Consoleに“common.on”と打ち込んでEnter
⇒ common.oldとcommon.newが束ねられている

```
> common.on
      common.old common.new
 [1,]          3          3
 [2,]          6          6
 [3,]          9          9
 [4,]         12         12
 [5,]         15         15
 [6,]         18         18
 [7,]         21         21
 [8,]         24         24
 [9,]         27         27
[10,]         30         30
[11,]         33         33
[12,]         36         36
```

plink関数に引き渡すデータの作成5

5. 最後に，今まで作成してきたデータをまとめる
   - その際，as.irt.pars関数を利用する

- as.irt.pars関数の使い方（今の場合）
  - pars.on <- as.irt.pars(x = para.on, common = common.on, cat = cat.on, poly.mod = pm.on)
    - x: 項目母数の推定値を指定する引数
      - 今の場合，2つのテストの項目母数（推定値）は”para.on”に入っているので，これを指定
    - common: 共通項目の情報を指定する引数
      - 今の場合，2つのテストに含まれる共通項目の情報は”common.on”に入っているので，これを指定
    - cat: 各項目のカテゴリ数を指定する引数
      - 今の場合，2つのテストの各項目のカテゴリ数は”cat.on”に入っているので，これを指定
    - poly.mod: 項目母数の推定時に利用した項目反応モデルを指定する引数
      - 今の場合，項目母数の推定時に利用した項目反応モデルの情報は”pm.on”に入っているので，これを指定

⇒ plink関数に引き渡すデータの作成が完了し，pars.onに

plink関数を利用し，等化係数を計算

- plink関数の使い方（今の場合，Mean & Sigma法を利用）
  - plink(x = pars.on, rescale = "MS", symmetric = TRUE)
    - x: as,irt.pars関数の出力した結果
      - 今の場合，as.irt.pars関数の出力した結果は”pars.on”に入っているので，これを指定
    - rescale: 等化を行う際に利用する等化係数を指定する引数
      - "MS"と指定した場合，Mean & Sigma法による等化が実行可能に
      - Mean & Sigma法の他に以下の等化法が利用可能
        - "MM": Mean & Mean法による等化を実行
        - “HB": Haebara法による等化を実行
        - "SL": Stocking-Lord法による等化を実行
    - symmetric: Haebara法やStocking-Lord法を実行する際に有効になる引数
      - “TRUE（はい）”と指定すれば，等化係数を求める際に，他のテスト版から得られた項目母数を規準集団の項目母数に等化した際の結果と，その逆の結果を考慮して，最終的な等化係数を決定する
  - out.on <- plink(x = pars.on, rescale = "MS", symmetric = TRUE)

⇒ テストoldを規準集団とした時の等化係数の推定値がout.onに

R Consoleに“summary(out.on)”と打ち込んでEnter
⇒ outo.onの中身の要約（summary）が出力
⇒ 4つの方法に基づく等化係数の推定値が得られる

```
>  summary(out.on)
-------  group2/group1*  -------
Linking Constants

                          A         B
Mean/Mean         1.217266 -0.557155
Mean/Sigma        1.168892 -0.515543
Haebara           1.106620 -0.457326
Stocking-Lord     1.114693 -0.479120
```

- 等化係数の推定値
  - A: Kに相当
  - B: Lに相当

plink関数で推定した等化係数を利用して等化を実行

- 等化係数の推定値を用いてテストnewの項目母数を等化
  - その際，link.pars関数を利用する
- link.pars関数の使い方（今の場合）
  - link.pars(x = out.on)
    - x: plink関数の出力した結果
      - 今の場合，plink関数の出力した結果は”out.on”に入っているのでこれを指定
  - para.equ <- link.pars(x = out.on)

⇒ テストnewの等化（equating）後の項目母数がpara.equに

R Consoleに“para.equ”と打ち込んでEnter
⇒　para.oldと等化後のpara.newがpara.equの中に

```
> para.equ
$group1
          [,1]     [,2]    [,3]
 [1,] 0.8704 -1.4507 0.1576
 [2,] 0.4628 -0.4070 0.1094
 [3,] 0.4416 -1.3349 0.1559
 [4,] 0.5448 -0.9017 0.1381
 [5,] 0.6200 -1.4865 0.2114
 [6,] 0.5730 -1.3210 0.1913
  ～中略～

$group2
           [,1]          [,2]    [,3]
 [1,] 0.4701889 -2.61487303 0.1751
 [2,] 0.6750838 -1.07614360 0.1165
 [3,] 0.3893431 -1.34557321 0.2087
 [4,] 1.2356146  0.04938250 0.2826
 [5,] 0.8332677 -0.71191686 0.2625
 [6,] 0.4995329 -1.51693278 0.2038
```

- para.equの中身
  - $group1: para.old
    - １列目：1.7×識別力母数
    - ２列目：困難度母数
    - ３列目：当て推量母数
  - $group2: 等化後のpara.new
    - １列目：1.7×識別力母数
    - ２列目：困難度母数
    - ３列目：当て推量母数

等化前と等化後の比較
⇒ para.newの値が変化している

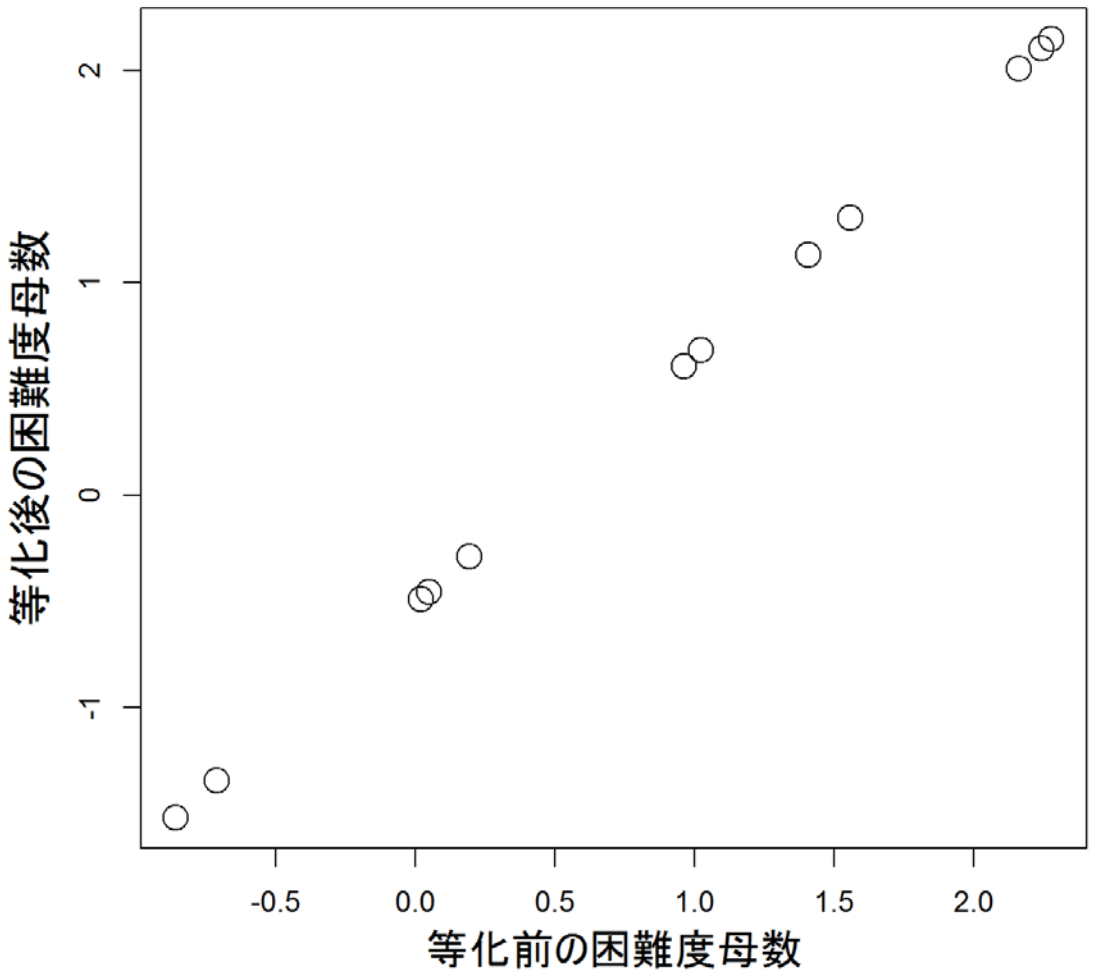
```
> para.on
[[1]]
        a.old    b.old   c.old
Item1  0.8704 -1.4507  0.1576
Item2  0.4628 -0.4070  0.1094
Item3  0.4416 -1.3349  0.1559
Item4  0.5448 -0.9017  0.1381
Item5  0.6200 -1.4865  0.2114
Item6  0.5730 -1.3210  0.1913
～中略～

[[2]]
        a.new    b.new   c.new
Item1  0.5496 -1.7960  0.1751
Item2  0.7891 -0.4796  0.1165
Item3  0.4551 -0.7101  0.2087
Item4  1.4443  0.4833  0.2826
Item5  0.9740 -0.1680  0.2625
Item6  0.5839 -0.8567  0.2038
Item7  0.8604 -0.4546  0.3224
```

```
> para.equ
$group1
          [,1]     [,2]    [,3]
[1,] 0.8704 -1.4507  0.1576
[2,] 0.4628 -0.4070  0.1094
[3,] 0.4416 -1.3349  0.1559
[4,] 0.5448 -0.9017  0.1381
[5,] 0.6200 -1.4865  0.2114
[6,] 0.5730 -1.3210  0.1913
  ～中略～

$group2
            [,1]          [,2]    [,3]
[1,] 0.4701889 -2.61487303  0.1751
[2,] 0.6750838 -1.07614360  0.1165
[3,] 0.3893431 -1.34557321  0.2087
[4,] 1.2356146  0.04938250  0.2826
[5,] 0.8332677 -0.71191686  0.2625
[6,] 0.4995329 -1.51693278  0.2038
[7,] 0.7360817  0.01583530  0.3224
```

テストnewの共通項目の困難度母数に関して散布図を作成
（x座標: 等化前のもの，y座標: 等化後のもの）
⇒　等化の前後で値が変化していることが見て取れる



最後に

IRTに基づきデータを分析する際の留意点

- 宇佐美先生，光永先生，登藤の発表
  - IRTに基づく分析の利点やその実行方法について解説

  ⇒　データが分析上の仮定を満たしていることが大前提
- IRTに基づく分析でよく利用される項目反応モデル
  - 候補となるモデルがたくさん存在する場合も
    - 二値型（1かO）のデータに対する1PLMや2PLM，3PLM

    ⇒　情報量規準を用いたモデル選択等が必要になる
  - データに対してかなり強い仮定を置いている
    - 局所独立性，1次元性，正答確率の単調増加性，…

    ⇒　これらの仮定が満たされているか必ず検討が必要
- IRTによる等化の実行
  - 適当な項目反応モデルを当てはめただけでは不十分

  ⇒　適切なテスト実施計画やデータ収集デザインが不可欠

引用文献

- Chen, W. H., & Thissen, D. (1997). Local dependence indexes for item pairs using item response theory. *Journal of Educational and Behavioral Statistics*, *22*(3), 265-289.
- Horn, J. L. (1965). A rationale and test for the number of factors in factor analysis. *Psychometrika*, *30*(2), 179-185.
- 喜岡恵子. (1991). 項目反応理論に基づく計算課題の難易度の推定. *教育心理学研究*, *39*(2), 204-213.
- 熊谷龍一. (2009). 初学者向けの項目反応理論分析プログラム EasyEstimation シリーズの開発. *日本テスト学会誌*, *5*(1), 107-118.
- R Core Team (2017). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.
- 酒井渉・野口裕之 (2015). 大学生を対象とした精神的健康度調査の共通尺度化による比較検討. *教育心理学研究*, *63*(2), 111-120.
- Shojima, K. (2014). Exametrika Ver. 5.3.
- Zimowski, M. F., Muraki, E., Mislevy, R., & Bock, R. D. (1996). BILOG-MG. *Multiple group IRT analysis and test maintenance for binary items*.

補足資料

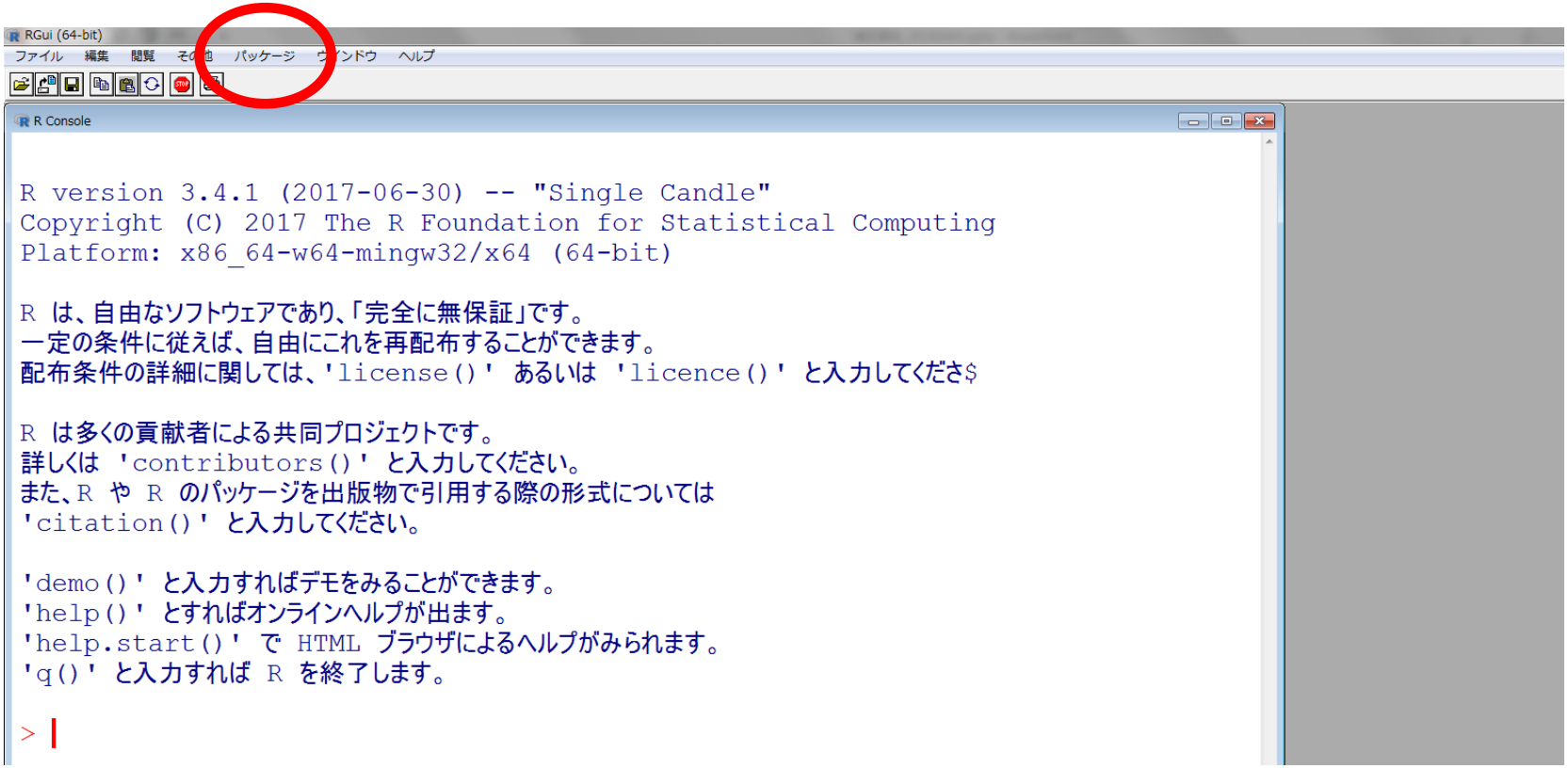
登藤　直弥
（筑波大学人間系）

Rにおけるパッケージのインストール方法等

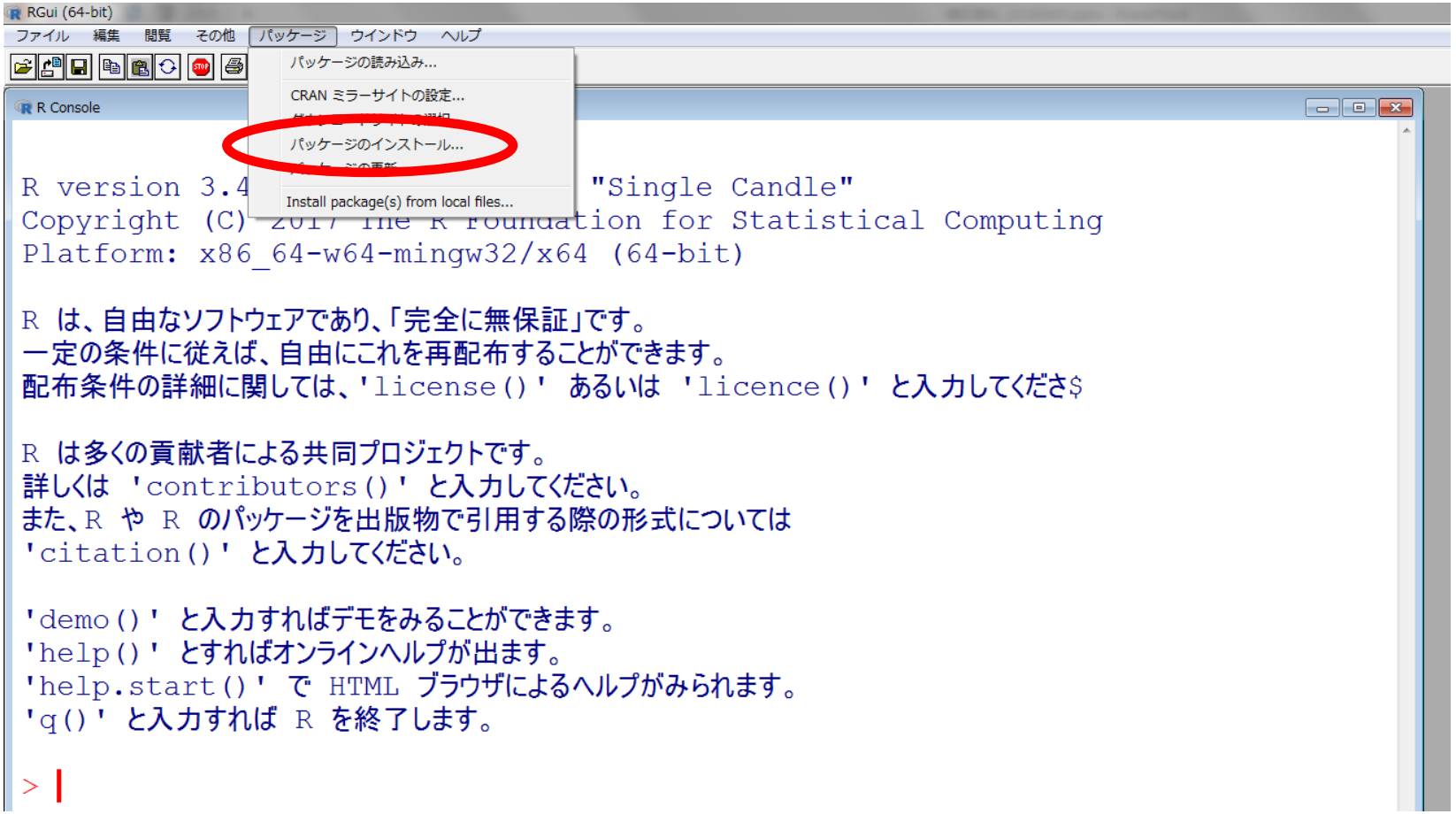
Rのパッケージとは

- Rをインストールしたばかりの初期状態
  ⇒　標準的な統計手法を利用することは可能
- よりアドバンスドな統計手法（Rの関数）
  ⇒　各種パッケージをPCにインストールすることで実行可能
- パッケージのインストール手順
  1. 「パッケージ」をクリック
  2. 「パッケージのインストール」をクリック
  3. 適当なミラーサイトを選ぶ
  4. インストールしたいパッケージを選択する

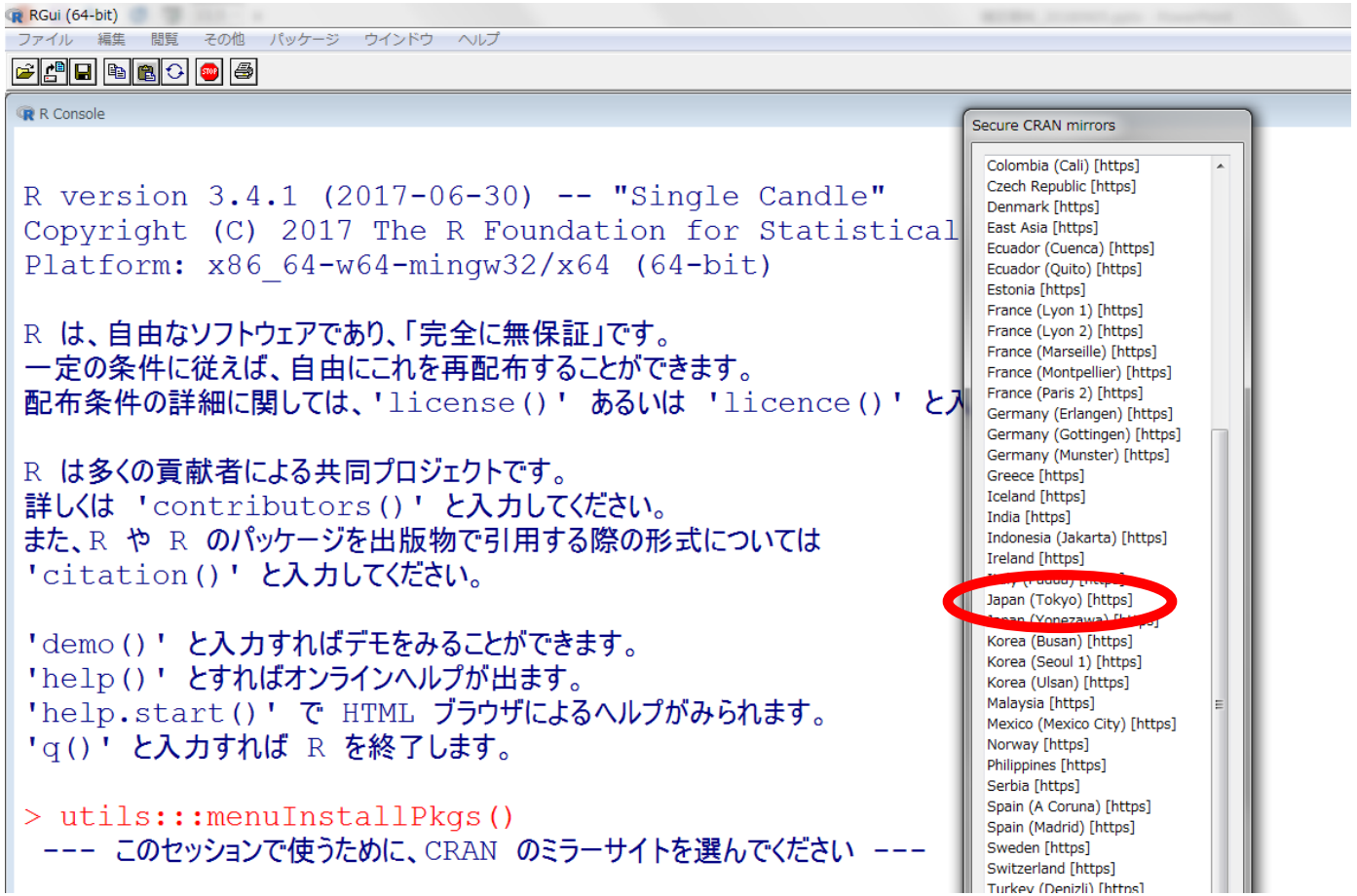
「パッケージ」をクリックする



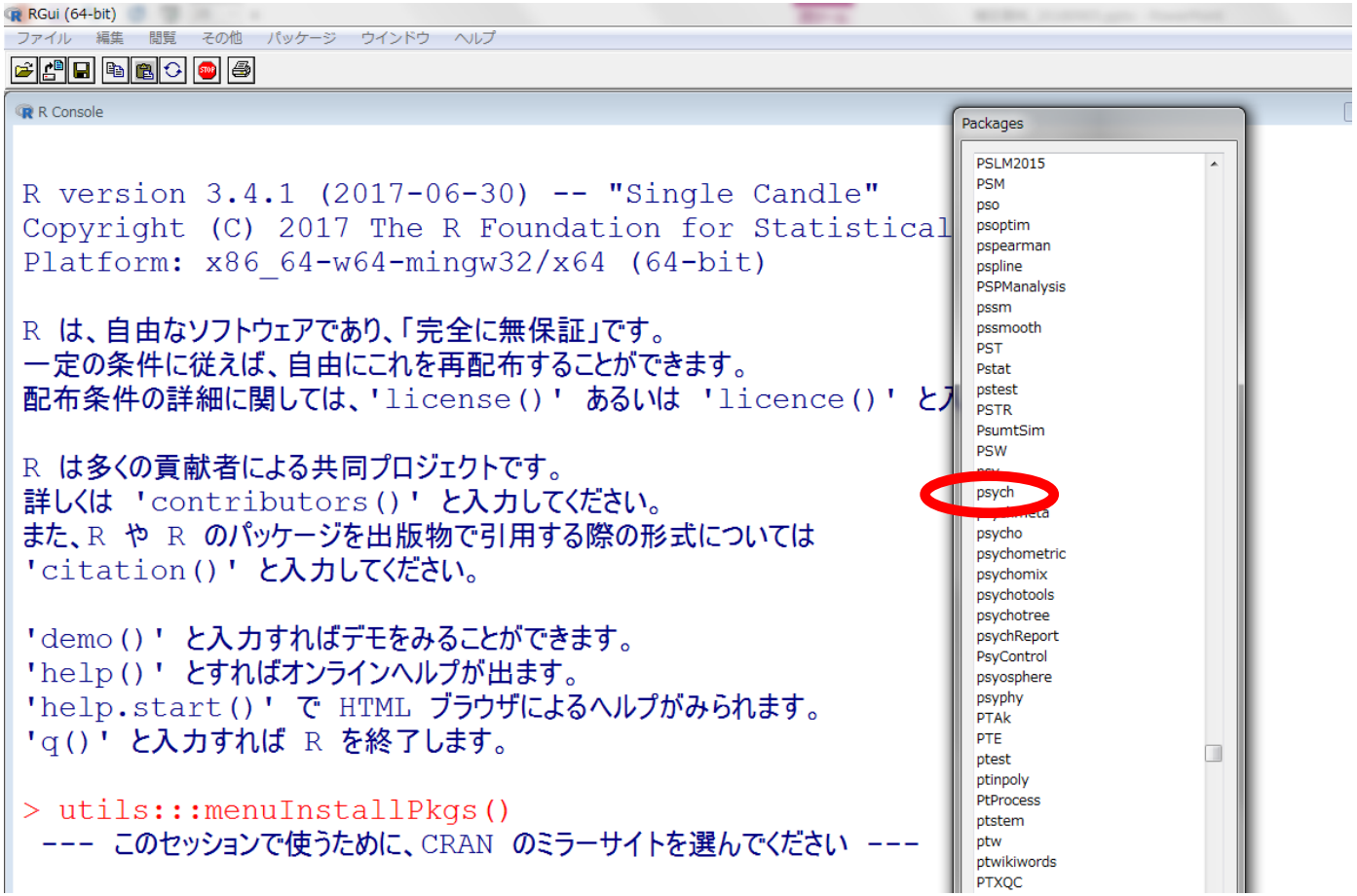
「パッケージのインストール」をクリック



適当なミラーサイトを選ぶ（e.g., Japan (Tokyo)）



インストールしたいパッケージを選ぶ



パッケージの使用方法

- 上記の手続きを実行
  ⇒　パッケージがPCへインストールされる
  ⇒　このままではパッケージ（Rの関数）を使用できない
- たとえば，ltmパッケージを使用する場合
  ⇒　R Consoleに

  library(ltm)

  と打ち込む

R Consoleに”library(ltm)”と打ち込む ⇒ ltmパッケージが使用可能に

```
> library(ltm)
 要求されたパッケージ MASS をロード中です
 要求されたパッケージ msm をロード中です
 要求されたパッケージ polycor をロード中です
 警告メッセージ:
1:  パッケージ 'ltm' はバージョン 3.4.4 の R の下で造られました
2:  パッケージ 'msm' はバージョン 3.4.4 の R の下で造られました
3:  パッケージ 'polycor' はバージョン 3.4.4 の R の下で造られました
```

分析用データの作成・読込方法

Rによる分析を始める前に

- Rが使用可能な状態に

  ⇒　実際にRを使用する際

  ⇒　分析用のデータを読み込む必要

- 以下では，
  - 分析用データの作成方法について解説
  - データのRへの読み込み方法について解説

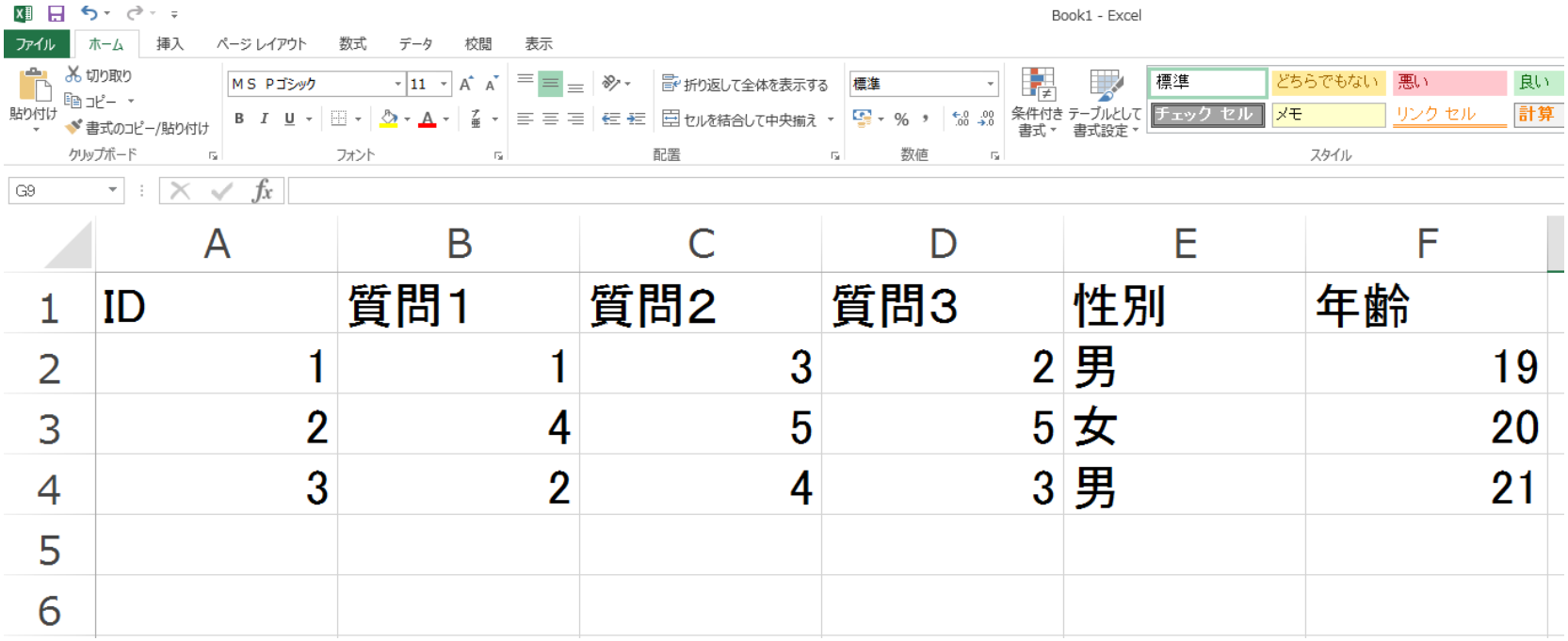
記入済みのテスト用紙，質問紙を回収

- Rで利用できるようにデータを打ち込まなければならない

  ⇒　Excelファイルにデータを記入することが多い

  ⇒　以下では，Excelファイルにデータを記入することを想定

回答をExcelファイルに記入する

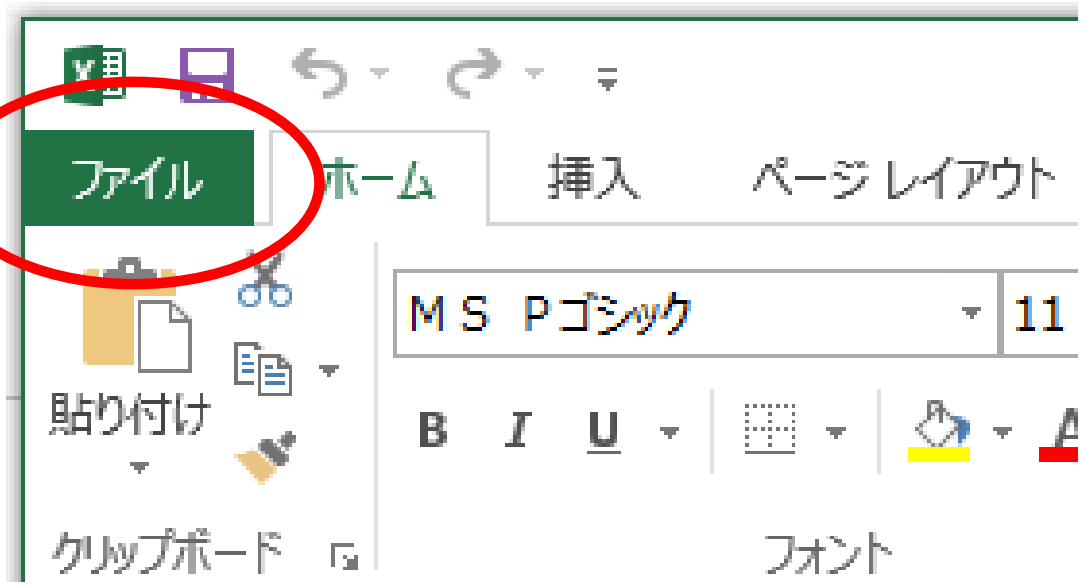


	A	B	C	D	E	F
1	ID	質問1	質問2	質問3	性別	年齢
2	1	1	3	2	男	19
3	2	4	5	5	女	20
4	3	2	4	3	男	21
5						
6						

ファイルを保存する

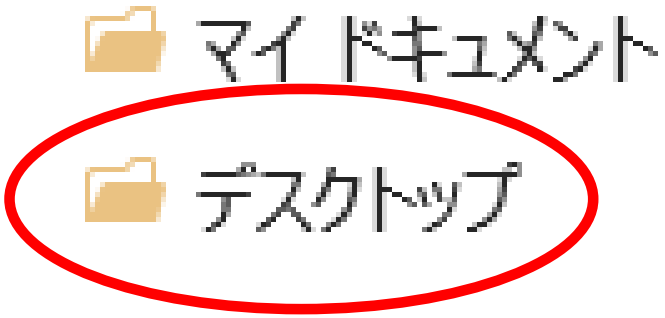
- Excelにデータを記入し終わったら，ファイルを保存する
  ⇒　CSVファイルとして保存する
  ⇒　Rに読み込む際に便利

「ファイル」を選択する

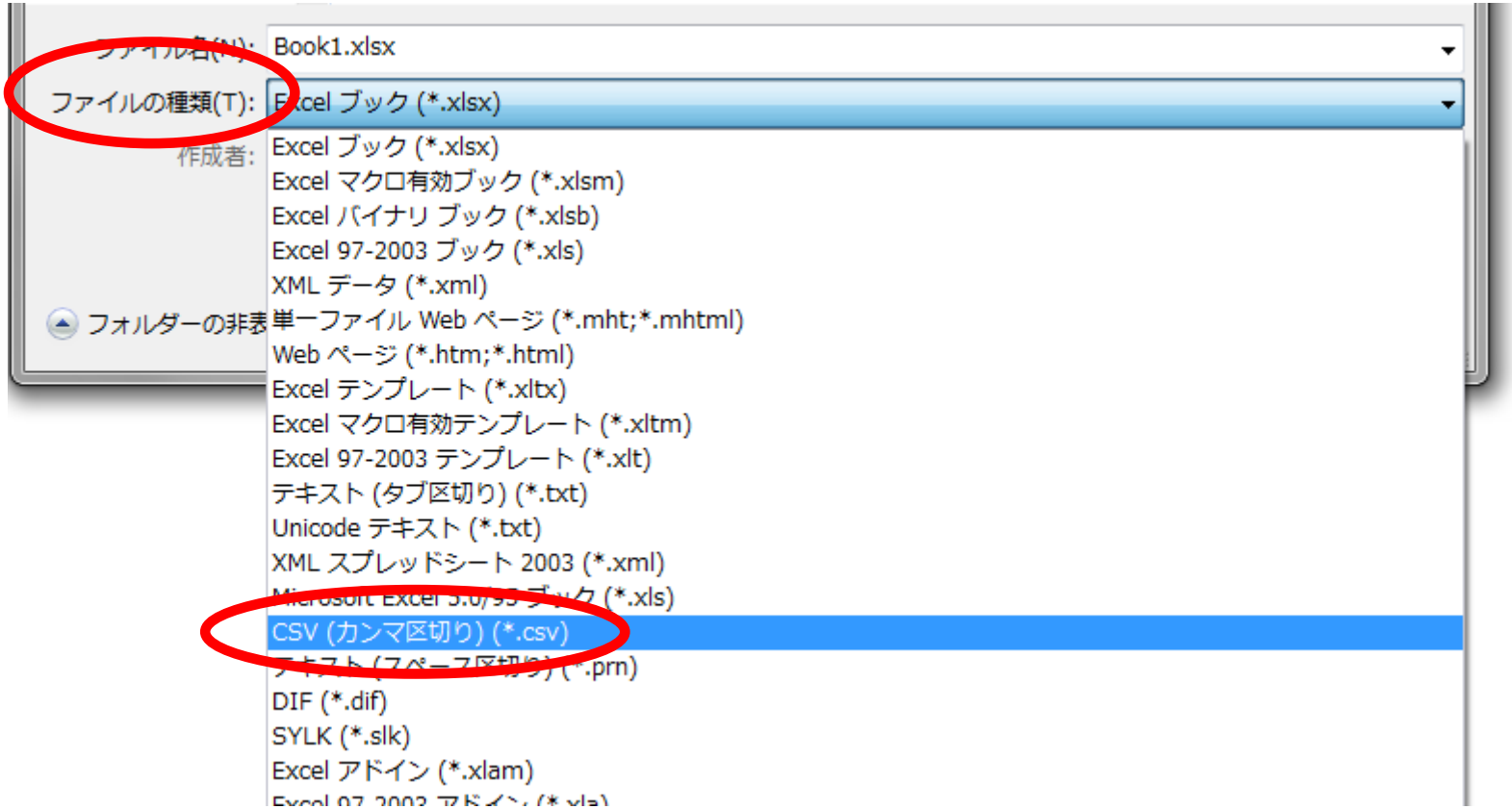


「名前を付けて保存」を選択する

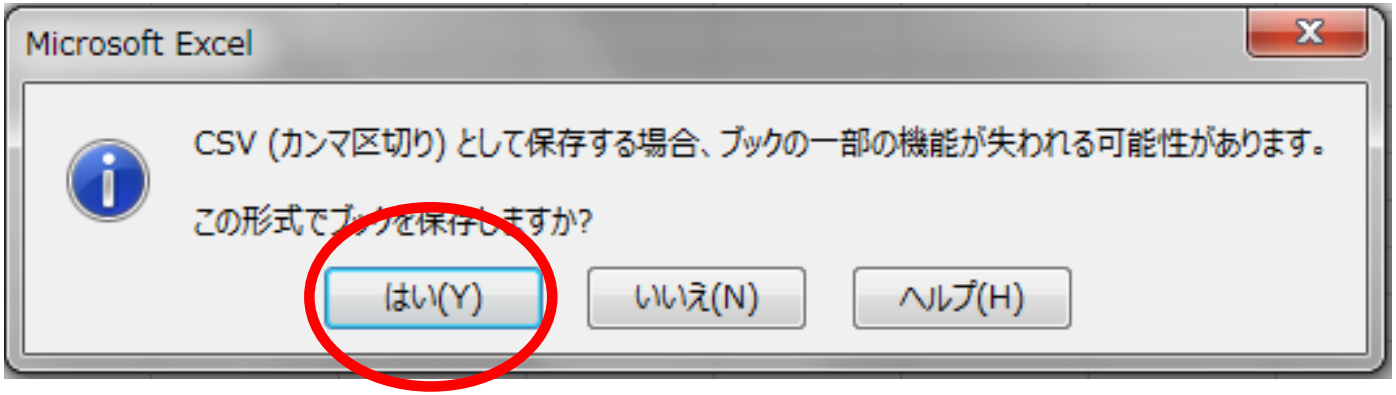
データファイルの保存先を選択する
⇒　たとえば，「デスクトップ」に保存



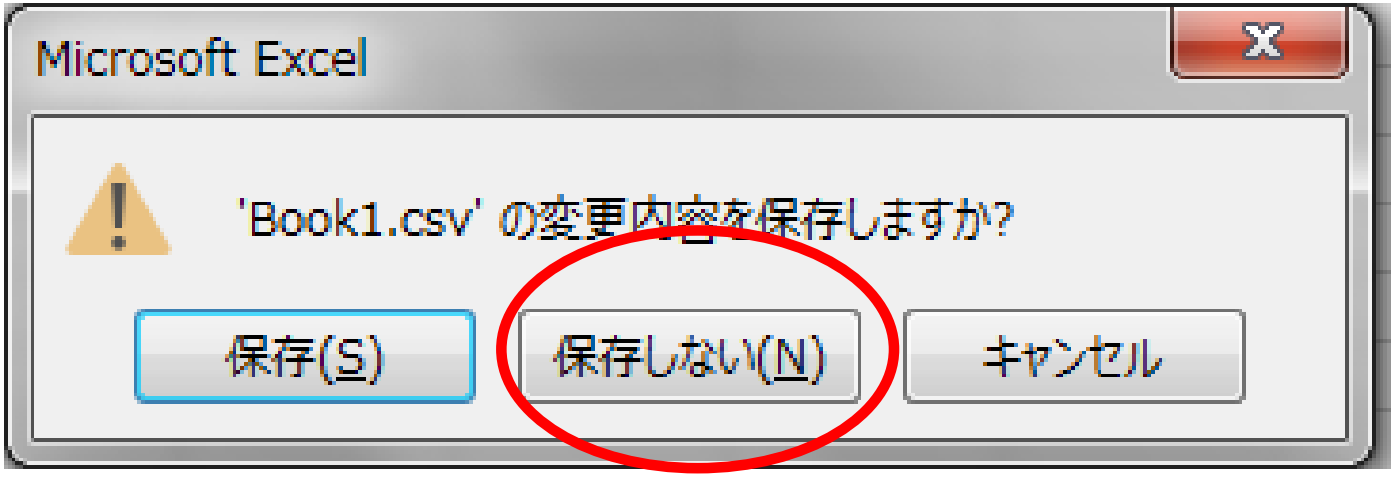
「ファイルの種類」→「CSV」として保存



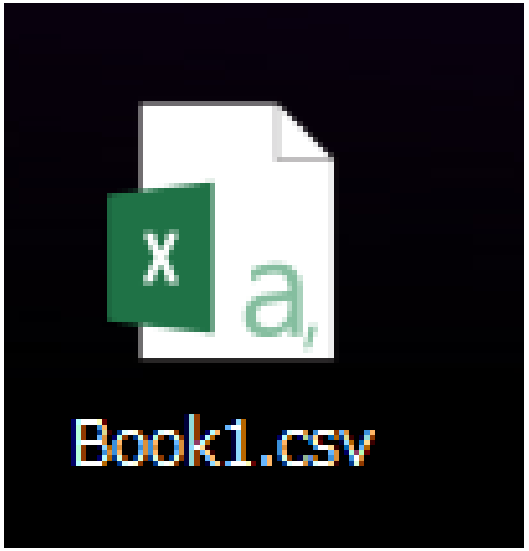
以下の画面が出るので「はい」を選択



ファイルを閉じるときに以下の画面が出る
⇒ 「保存しない」を選択



デスクトップにCSVファイルが作成



CSVファイルに関する諸注意

- CSVファイルで以下のこと実行し，保存する
  - 適当な行，列，またはセルに色をつける
  - 文字に色をつける
  - 列や行の幅を変える
  - …

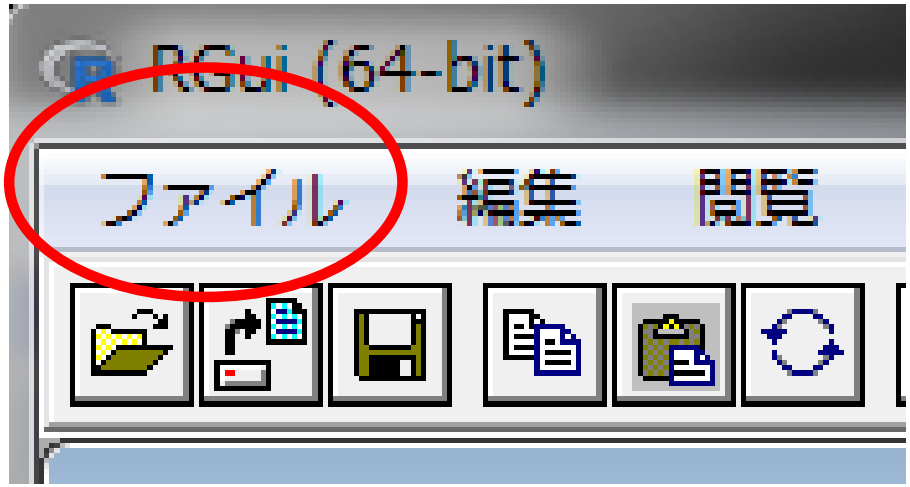
  ⇒　変更が反映されません
- また，以下のことを実行して保存する
  - 数字を四捨五入して表示

  ⇒　四捨五入する前の元データが失われます

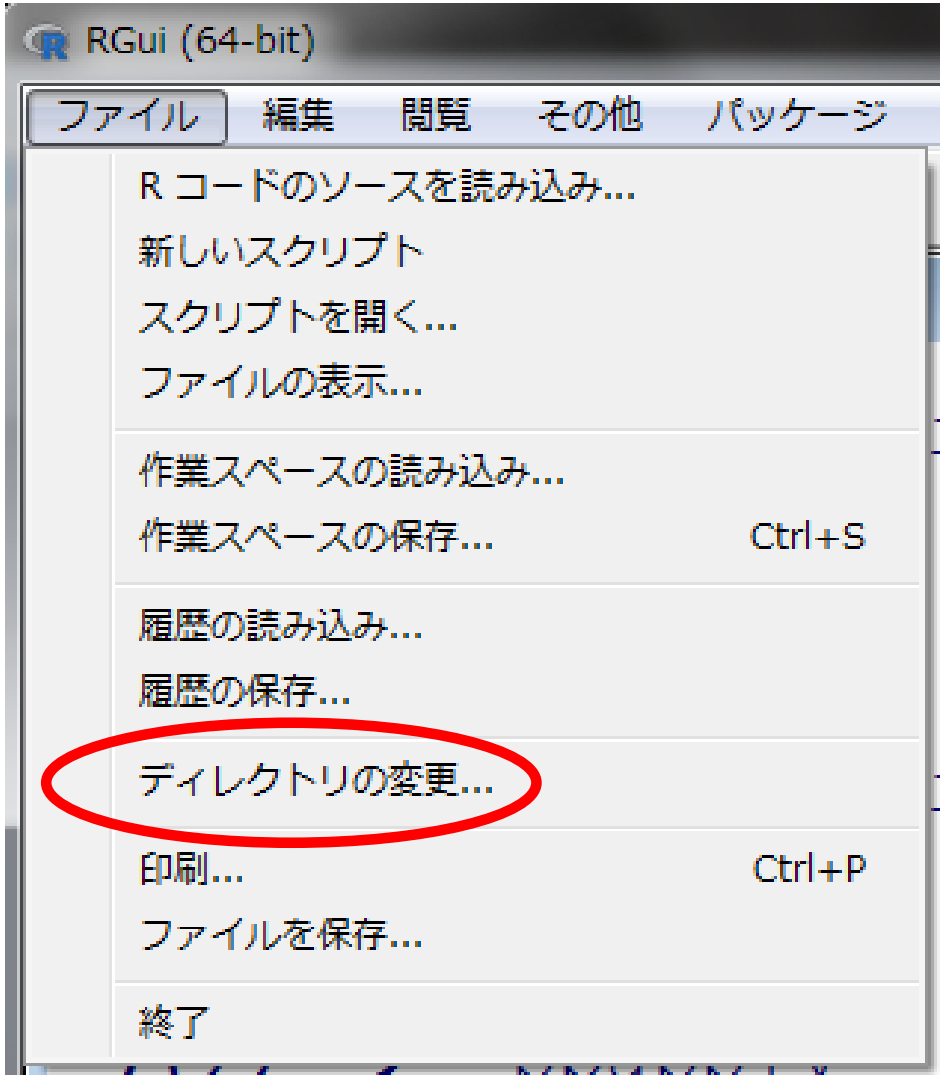
作成したCSVファイルをRに読み込む

- 分析用データのCSVファイルが完成
  ⇒　Rに読み込む
- RへのCSVファイルの読み込み
  ⇒　read.csvという関数を使用
- read.csv関数の使い方（今の場合）
  - read.csv(file = "Book1.csv", header = TRUE, row.names = 1)
    - file: 回答データの入っているCSVファイルの名前を指定する引数
      - 今の場合，回答データは"Book1.csv"に入っているので，これを指定する
      - “file”のように，関数の中で指定する値のことを「引数」という
    - header: CSVファイルの１行目に変数名が記載されているか否か
      - ”TRUE（はい）”と指定することでCSVファイルの１行目が変数名と認識される
    - row.names: 受験者IDが何列目に記載されているかを指定する引数
      - ”1”と指定することでCSVファイルの１列目が受験者IDと認識される
  - DATA <- read.csv(file = "Book1.csv", header = TRUE, row.names = 1)
  ⇒　読み込んだ回答データ（DATA）がDATAに

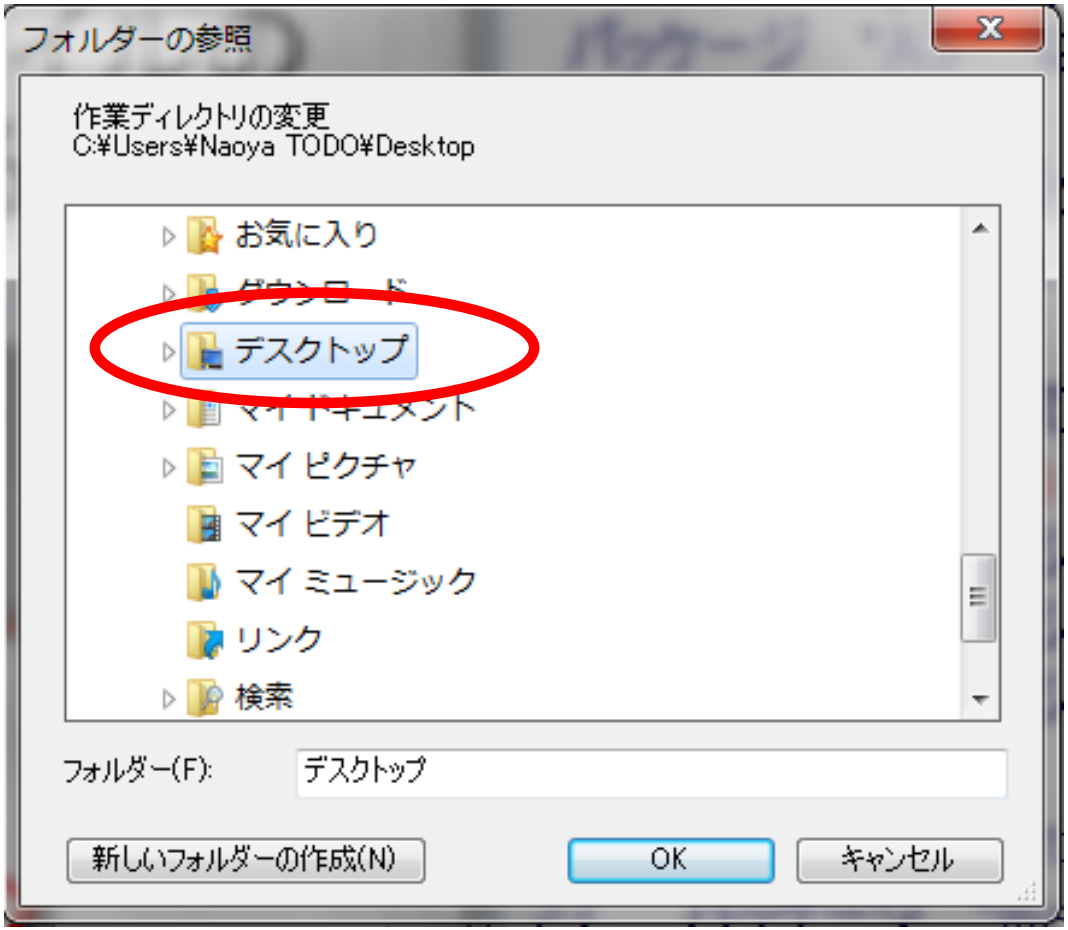
「ファイル」をクリック



「ディレクトリの変更」をクリック



CSVファイルのある場所を指定する

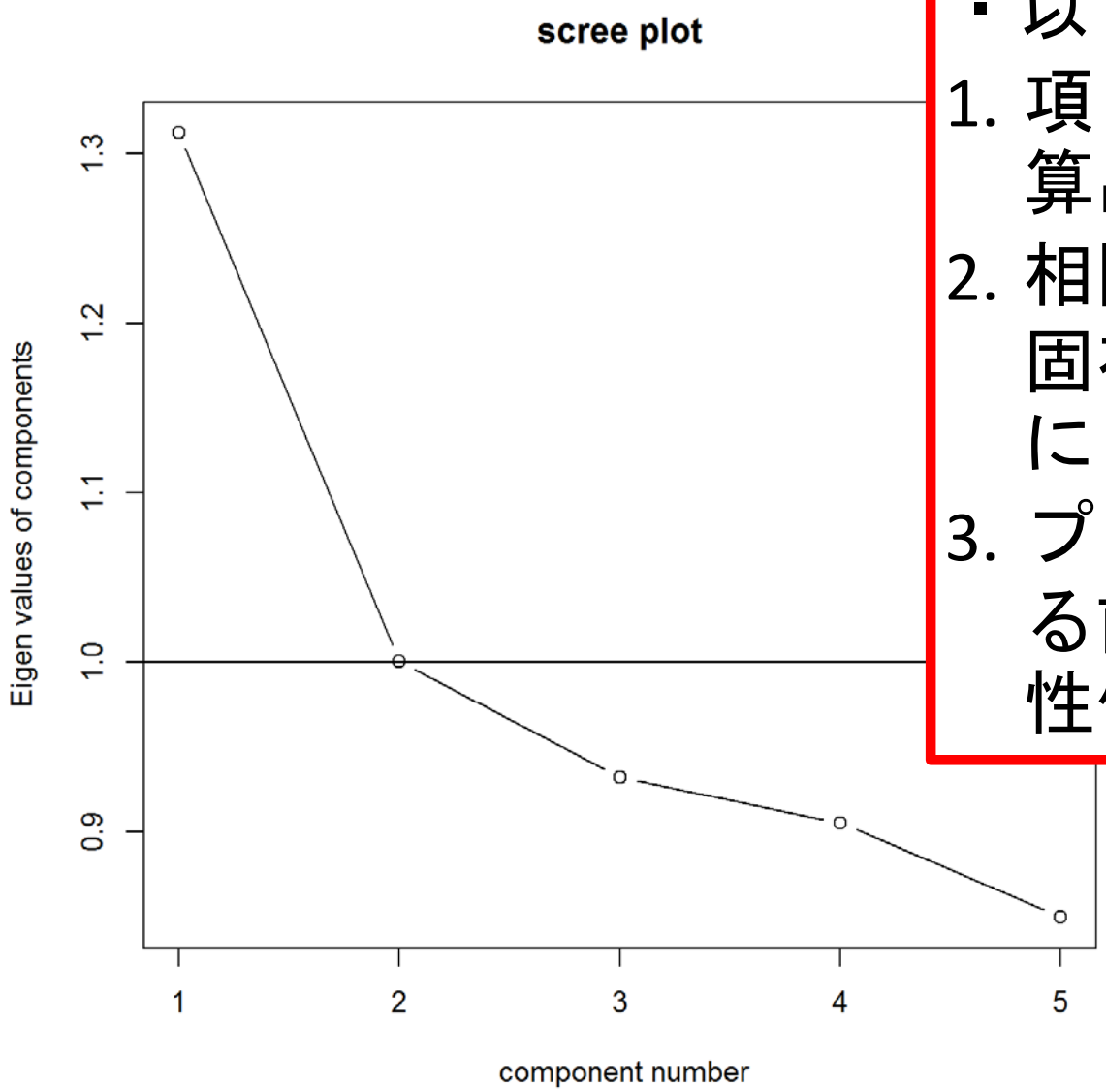


先のコマンドを実施する
⇒　R Consoleに”DATA”と打ち込んでEnter
⇒　読み込んだデータが表示される

```
> DATA <- read.csv(file = "Book1.csv", header = TRUE, row.names = 1)
> DATA
  質問1 質問2 質問3 性別 年齢
1     1     3     2   男   19
2     4     5     5   女   20
3     2     4     3   男   21
```

スクリープロットおよび平行分析について

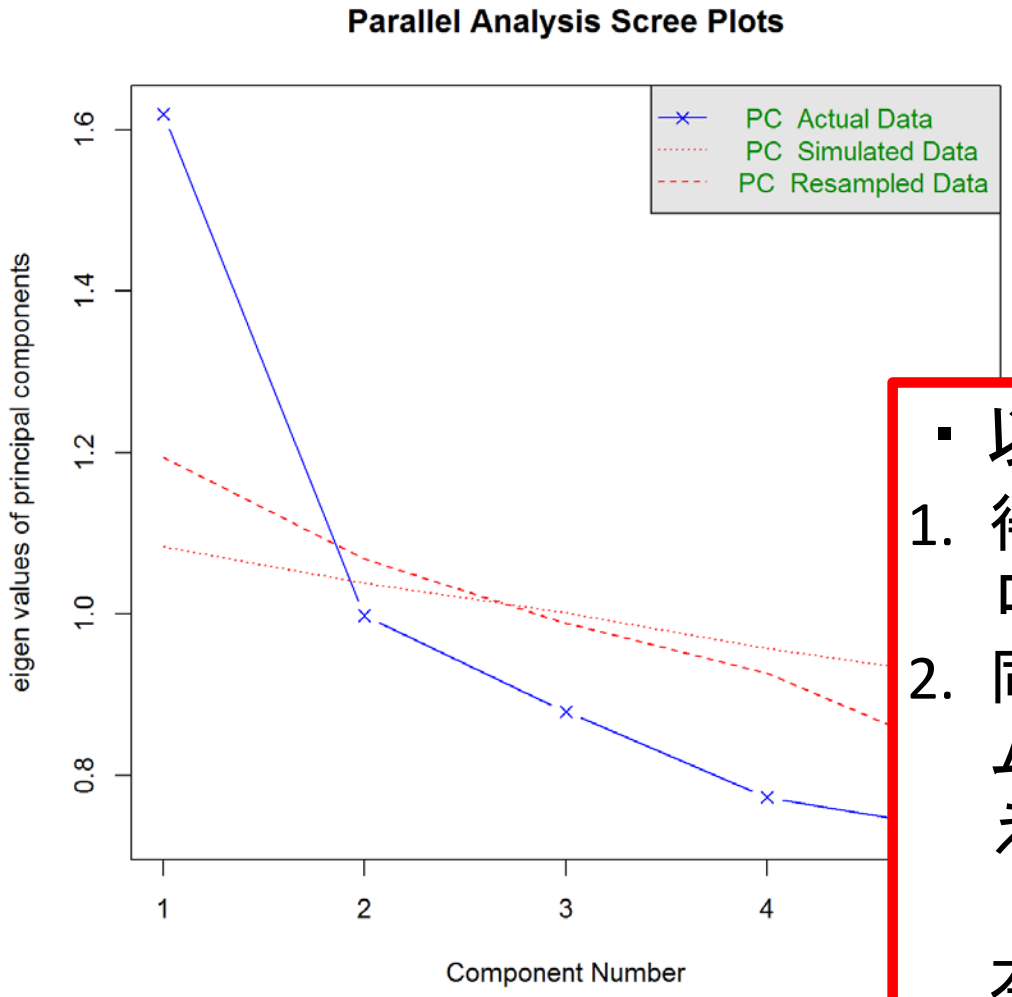
スクリープロットとは



・以下の手順で求める

1. 項目得点間の相関行列を算出
2. 相関行列を固有値分解し，固有値を得て，大きい順にプロットする
3. プロットがなだらかになる前の固有値数を潜在特性値の次元数とする

平行分析（Horn, 1965）とは



- 以下の手順で実行する

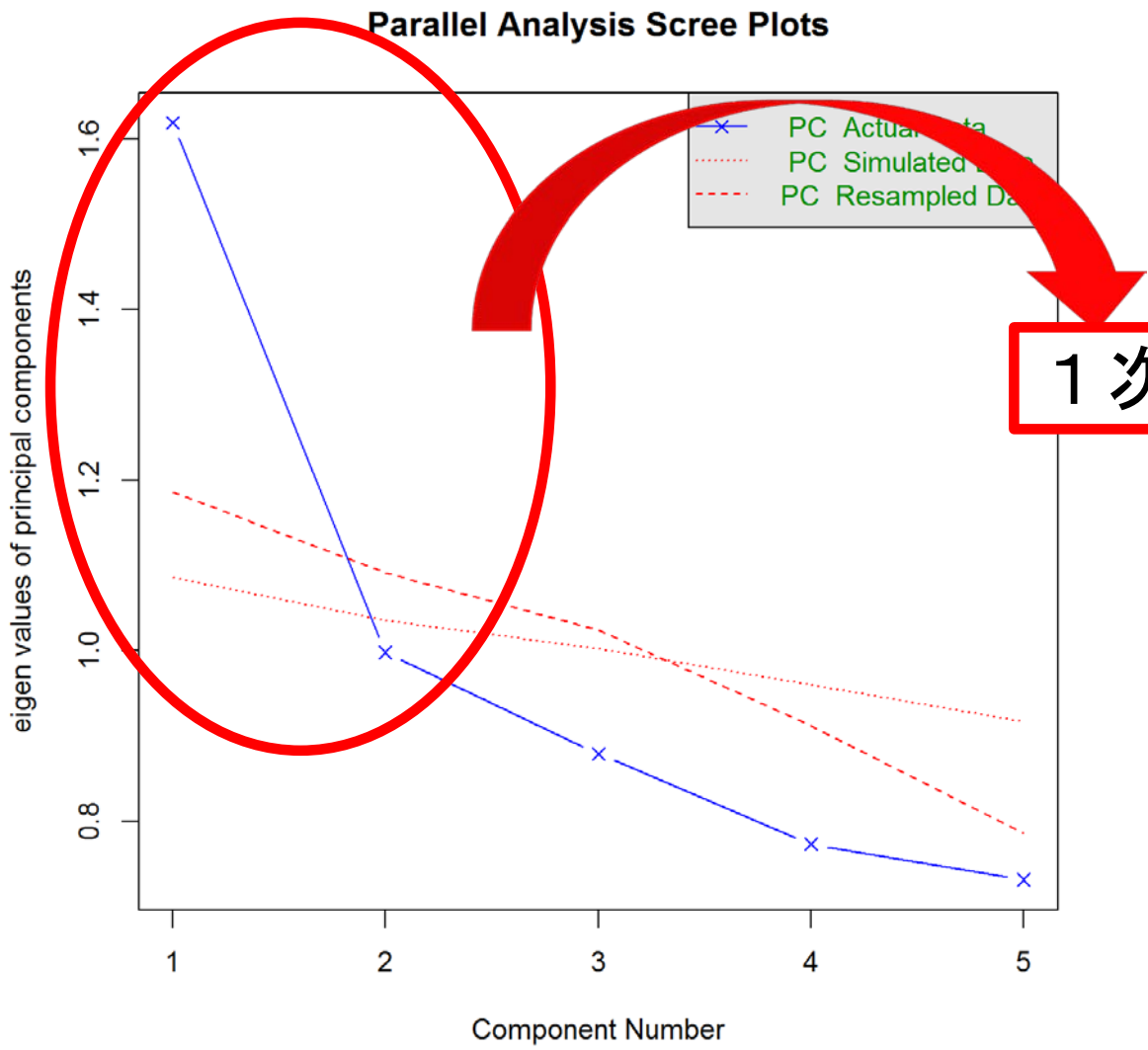
1. 得られたデータのスクリープロットを作成（左図の青実線）
2. 同じ受験者数と項目数のランダムデータを複数発生させ，たとえば，平均的なスクリープロットを作成（今の場合，左図の２本の赤点線）
3. １と２を比較して，２よりも上に位置する１の固有値の数を潜在特性値の次元数とする

Rでのスクリープロットの作成と平行分析の実行方法

- Rでのスクリープロットの作成と平行分析の実行
  - fa.parallel関数（平行分析を実行する）でまとめて実行可能

  ⇒　fa.parallel関数はpsychパッケージの中に入っている関数
- fa.parallel関数の使い方（テストデータの例の場合）
  - fa.parallel(x = TEST[, 1:5], fa = "pc", cor = "tet")
    - x: テストデータを指定する引数
    - fa: 固有値分解の対象となる行列を指定する引数
      - 項目得点間の相関行列を固有値分解するので"pc"と指定
    - cor: 項目得点間の相関係数として何を計算するかを指定する引数
      - "tet"とすると，各項目得点（二値）の背後に連続的な潜在変数（正規分布に従う）があると仮定し，この潜在変数同士の相関（四分相関）を計算する

先のプログラムを実行する
⇒　平行分析の結果が表示される



１次元性が示唆された