



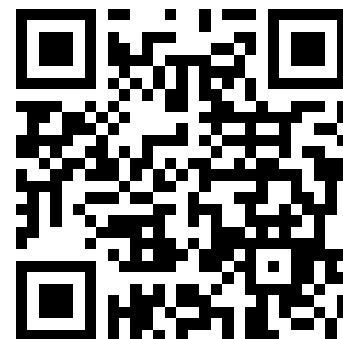
研究のオープン化の観点から
私の講演は**撮影・録音はご自由に.**
※ただし, 他人へのご配慮をお願いします

日本認知・行動療法学会
第43回大会 @朱鷺メッセ

質問紙における回答バイアス統制の ためのベイズ統計モデリング

専修大学大学院文学研究科 修士2年 北條大樹

連絡先: dhojou@psy.senshu-u.ac.jp
本日の資料: <https://dastatis.github.io/index.html>



自己紹介



- ◆ 名前：北條大樹（ほうじょう だいき）
- ◆ 出身：新潟
- ◆ 所属：専修大学大学院 文学研究科 修士2年
- ◆ 専門：心理統計学・ベイズ統計学
- ◆ 研究内容：質問紙における回答傾向バイアスの補正

今心理学で大流行!?!のベイズ統計モデリング

ベイズ統計で 実践モデリング

認知モデルのトレーニング

M. D. リー &
E.-J. ワーゲンメイカーズ 著

井関 龍太 訳 岡田 謙介 解説

北大路書房

- ここにいるということは...
 - **ベイズに興味がある**
 - **ベイズを使っている**という方々だと思います
- この本には、**ベイズの基礎から、心理学実験にベイズ統計モデリングを適用**した例がわかりやすく記載されています
- あ、一昨日発売されたばかりですので、みなさまぜひ!!!
(書籍販売コーナーにあります)

でも、私、実験屋でなく、調査屋だからなあ...

- 私の講演は、**調査研究を行う臨床心理学的研究者を対象**にしたベイズ統計の2つの使い方についてお話いたします

ベイズの2つの使い方

• ベイズ統計データ分析

t検定・分散分析などといった分析をベイズ推定で行う

イメージ：既存の“型”にベイズを適用し、解釈をやすくする

• ベイズ統計モデリング

確率分布と数式を使って、データ生成メカニズムを表現

イメージ：データがどのように得られたか考え、適合する枠組みを作成

本日のお品書き

1. 質問紙って何? 調査って何?
2. 調査研究のための (ベイズ) 統計データ分析
 - IRTについて(Item Response Theory; 項目反応理論)
 - IRTをベイズで行う利点
3. 調査研究のための ベイズ統計モデリング
 - 回答スタイルバイアスについて
 - 係留ビネット法 (Anchoring Vignettes method)
 - ベイズ多次元IRTモデル

1. 質問紙って何? 調査って何?

1. はじめに

Q. なぜ、我々は質問紙を取るのか？

A. 調査・測定を行うため

Q. では、なぜ調査・測定を行う？

A. 研究仮説を検証するため

A. 回答者の能力を測るため



1. はじめに

Q. なぜ、我々は質問紙を取るのか？

A. 調査・測定を行うため

Q. では、なぜ調査・測定を行う？

A. 研究仮説を検証するため

A. 回答者の能力を測るため

あなたの質問紙 (の取り方) で
本当に検証・測定したいものを
取得できていますか？



1. 一般的な質問紙調査

- 項目の**素点**や**合計**・**平均**をそのまま利用することが多い

ID	Q1	Q2	Q3	合計	平均
1	1	4	3	8	2.67
2	4	4	3	11	3.67
3	1	2	2	5	1.67

※素点 -- 項目のそのままの得点のこと



1. これまでの調査イメージ

構成概念
(測定したいもの)

反映



一般的な調査

素点 → 単純集計(平均)での解釈

1. 質問紙調査の問題例 1

加藤・山田・川端 (2014)
Rによる項目反応理論

・集団・標本依存性

- ・片方の集団が優秀である場合は、調査得点を同様に解釈して使用することができない



スーパー寺子屋 高校
ある調査における
平均得点 = 15点



自由奔放 高校
ある調査における
平均得点 = 15点

この2つの高校間の平均得点の解釈は一緒ではない

・項目依存性

- ・ 同じ調査内容(e.g. 抑うつ)について尋ねていても、調査項目が異なる場合、調査得点を同様に解釈できない。

・ 最近、あなたは深く考え込んでしまうことがある。

・ 最近、あなたは深く考え込んでしまうことがない。

- | | | | | |
|------------|-------------|---------|-----------|----------|
| 1 | 2 | 3 | 4 | 5 |
| とてもあてはまらない | だいたいあてはまらない | どちらでもない | だいたいあてはまる | とてもあてはまる |

ただの逆転項目のはずなのに
逆転していないことが報告

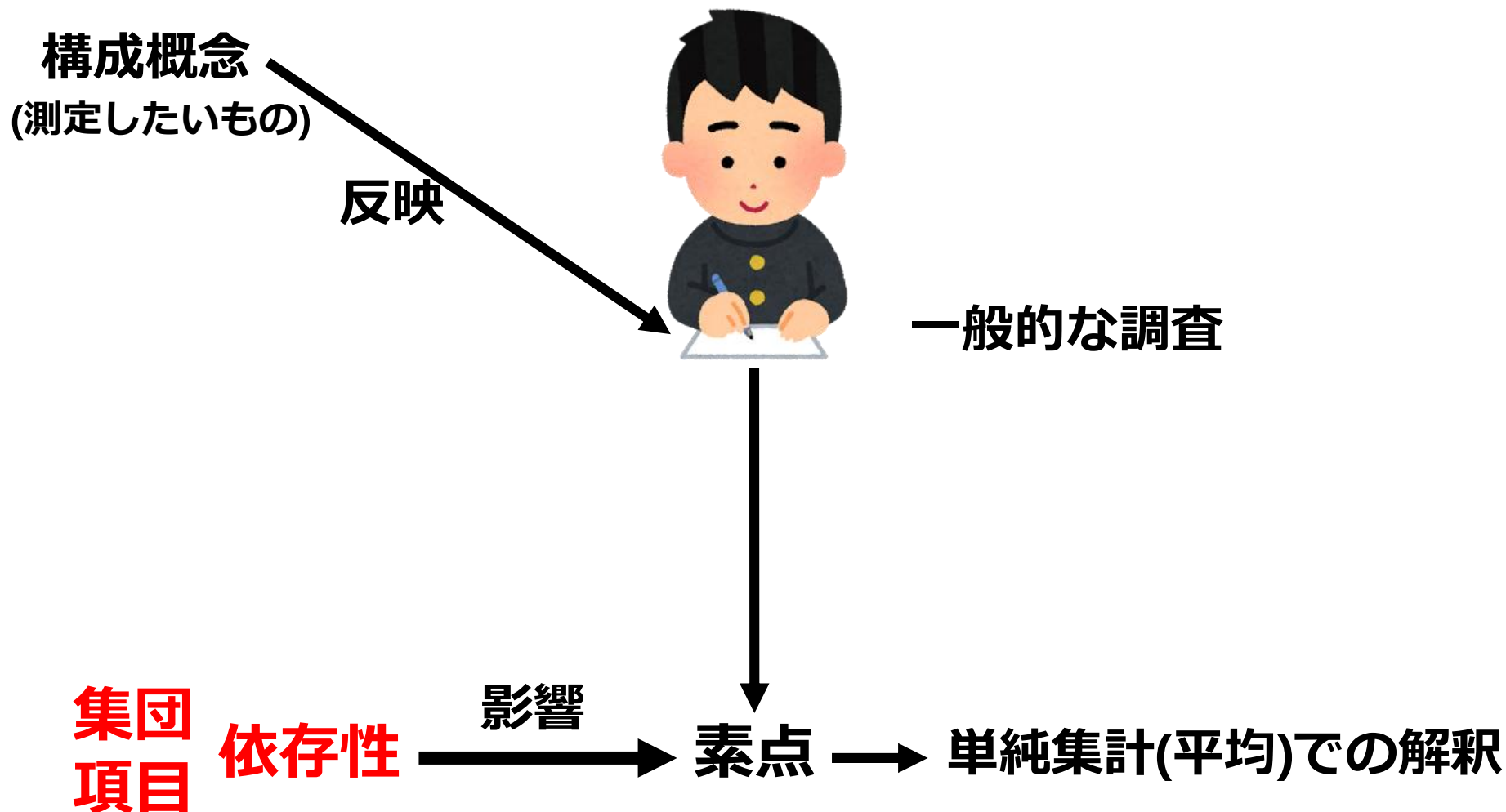
(Harpole et al. 2014; Rodebaugh et al. 2004)

素点のみの解釈では
このような問題点がある

2. 調査研究のための (ベイズ) 統計データ分析

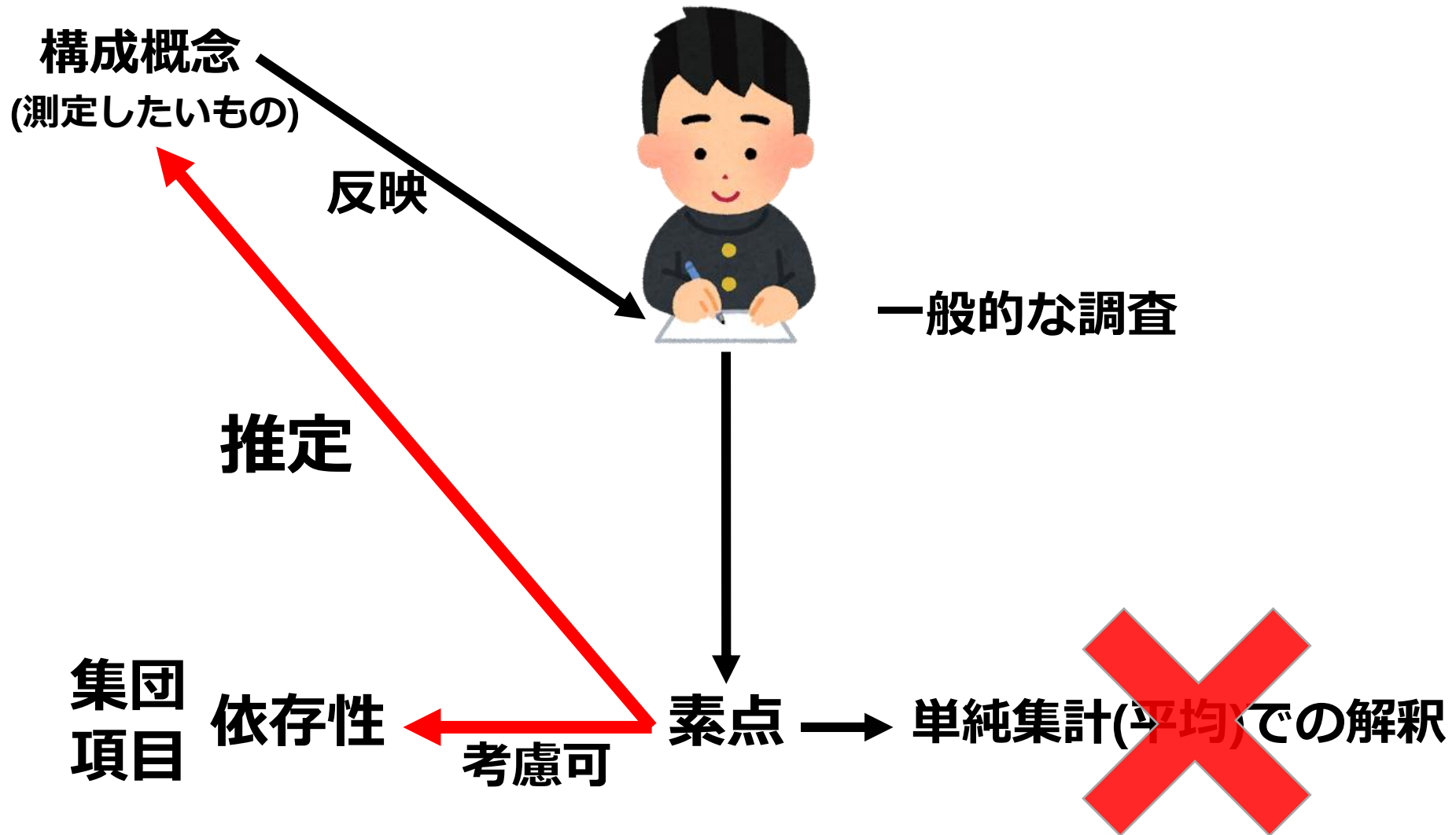
IRTについて(Item Response Theory; 項目反応理論)
IRTをベイズで行う利点

これまでの調査イメージ (再掲)



集団や項目の特性に依存する

これから話すことのイメージ

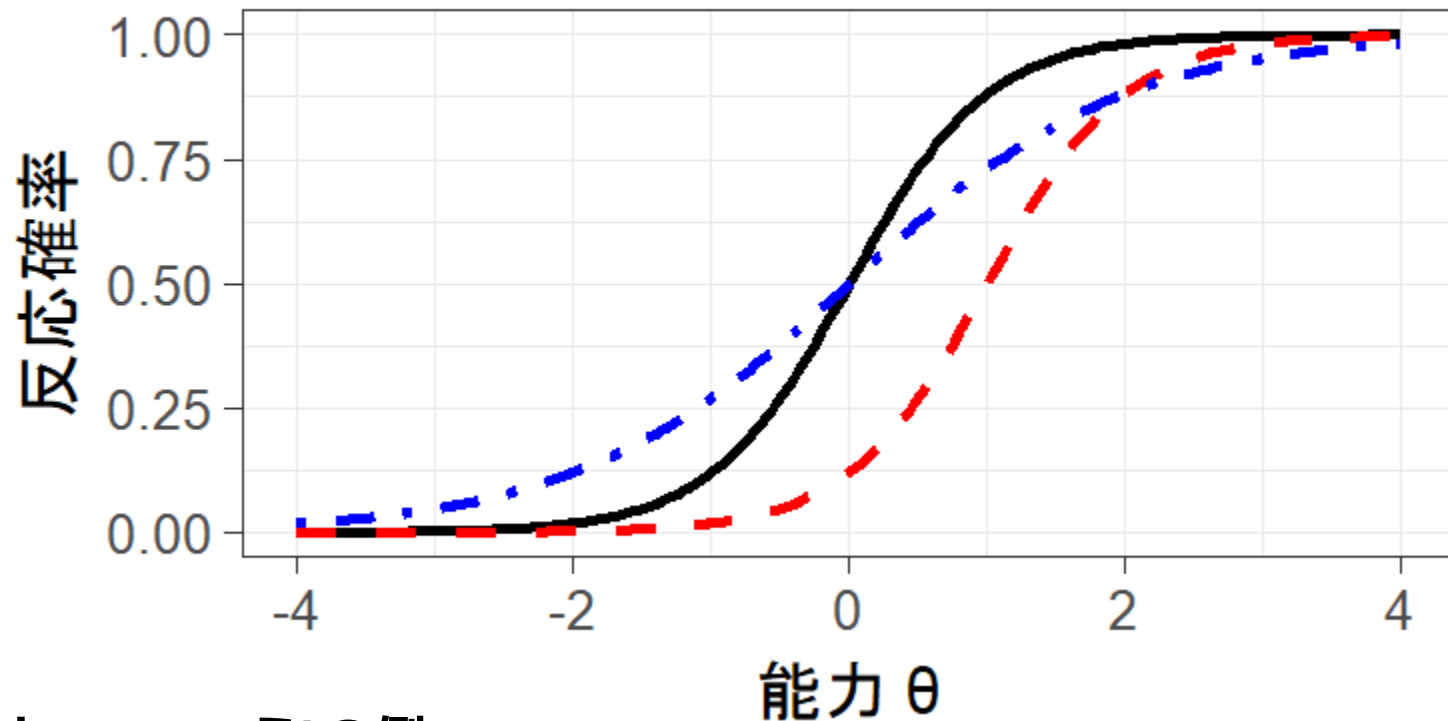


集団・項目の特徴を考慮し、**能力**を把握する

2. 先の問題に対処するためのIRT

- IRT (Item Response Theory; 項目反応理論)
 - 項目回答から **回答者の能力** と **項目の特徴** を推定可能な理論
 - IRTを応用することで回答者集団や項目特性に依存しない回答者の能力を測ることが可能になる

項目特性曲線(1つの線は1つの項目を表す)



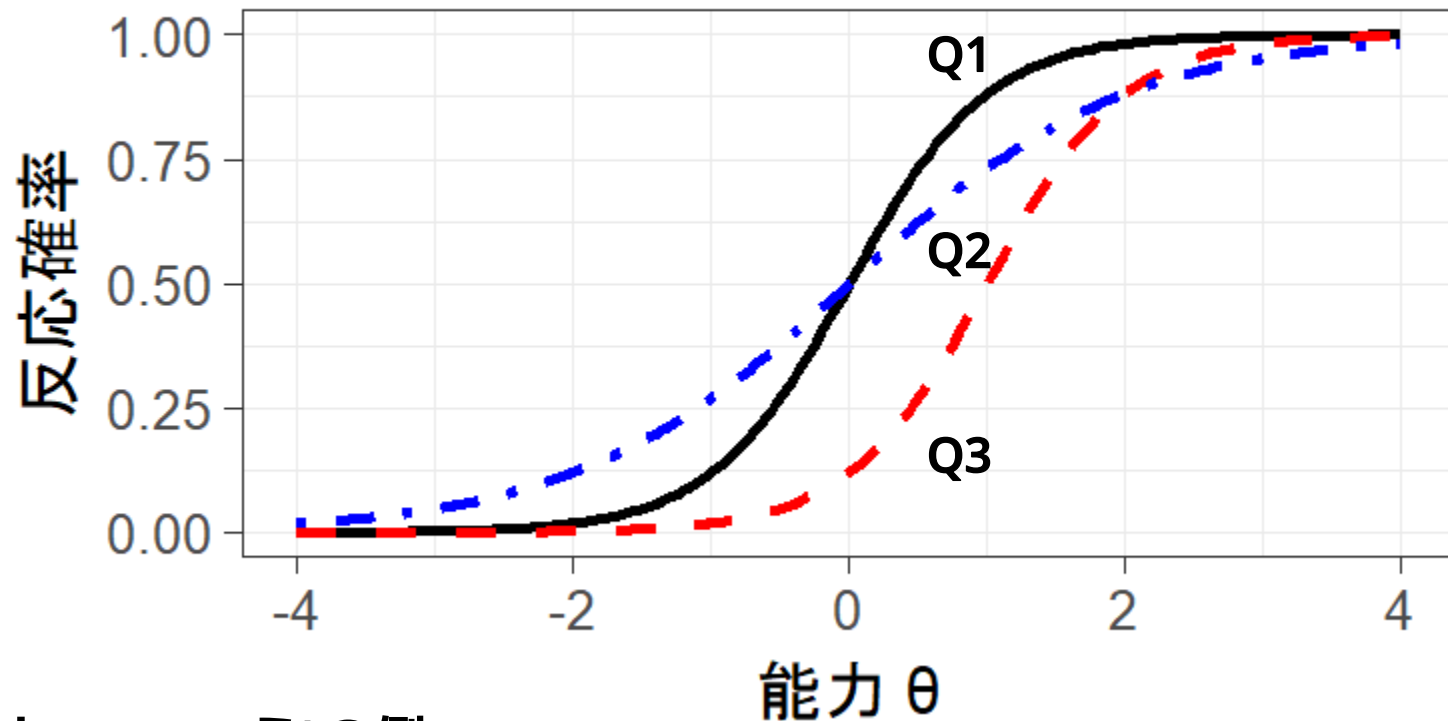
※2値(はい・いいえ)の例

2. IRTとは? (2パラメータロジスティックモデル)

ID	Q1	Q2	Q3
1	1	0	0
2	0	1	1
3	0	0	1

項目特性を推定し、
回答者の能力を推定

項目特性曲線(1つの線は1つの項目を表す)



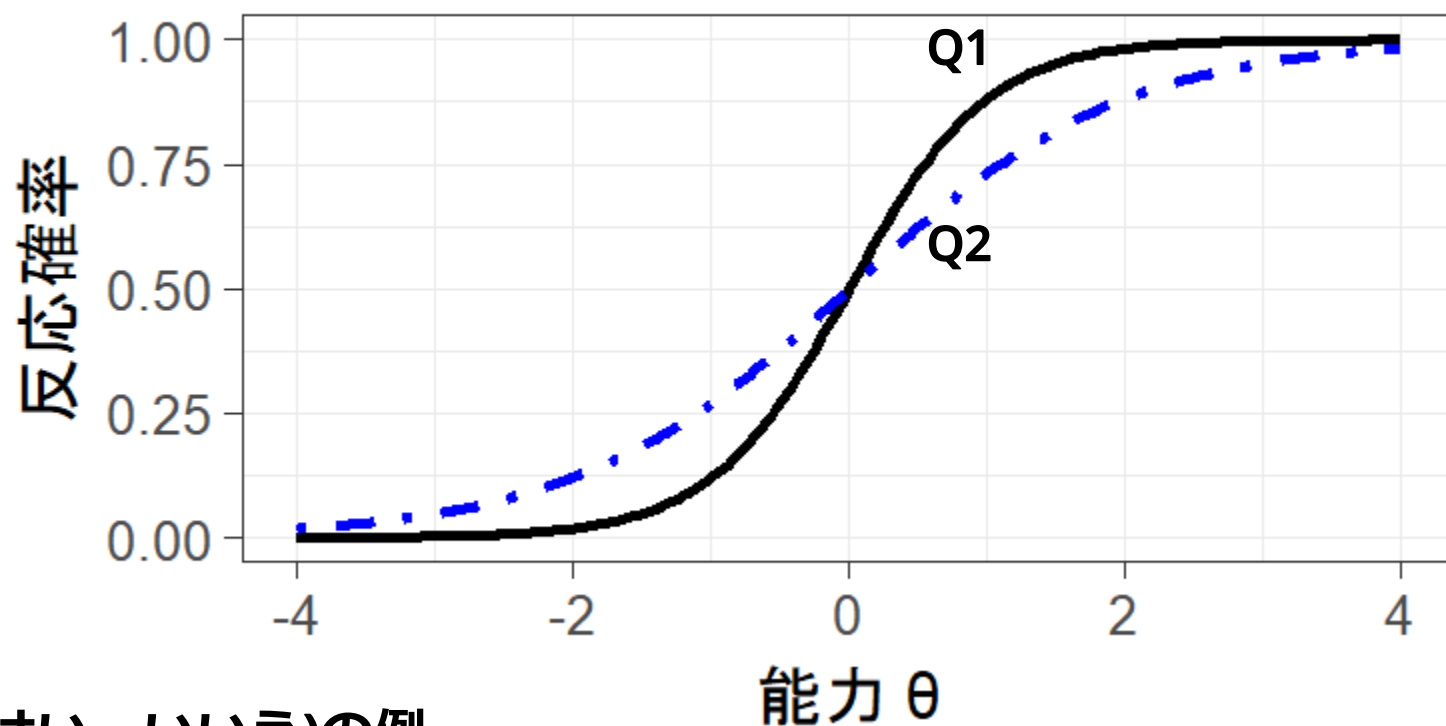
※2値(はい・いいえ)の例

2. 項目の何がわかるのか? (識別力)

ID	Q1	Q2	Q3
1	1	0	0
2	0	1	1
3	0	0	1

項目特性を推定し、
回答者の能力を推定

どれだけ 識別できる 項目かを検討できる(識別力)



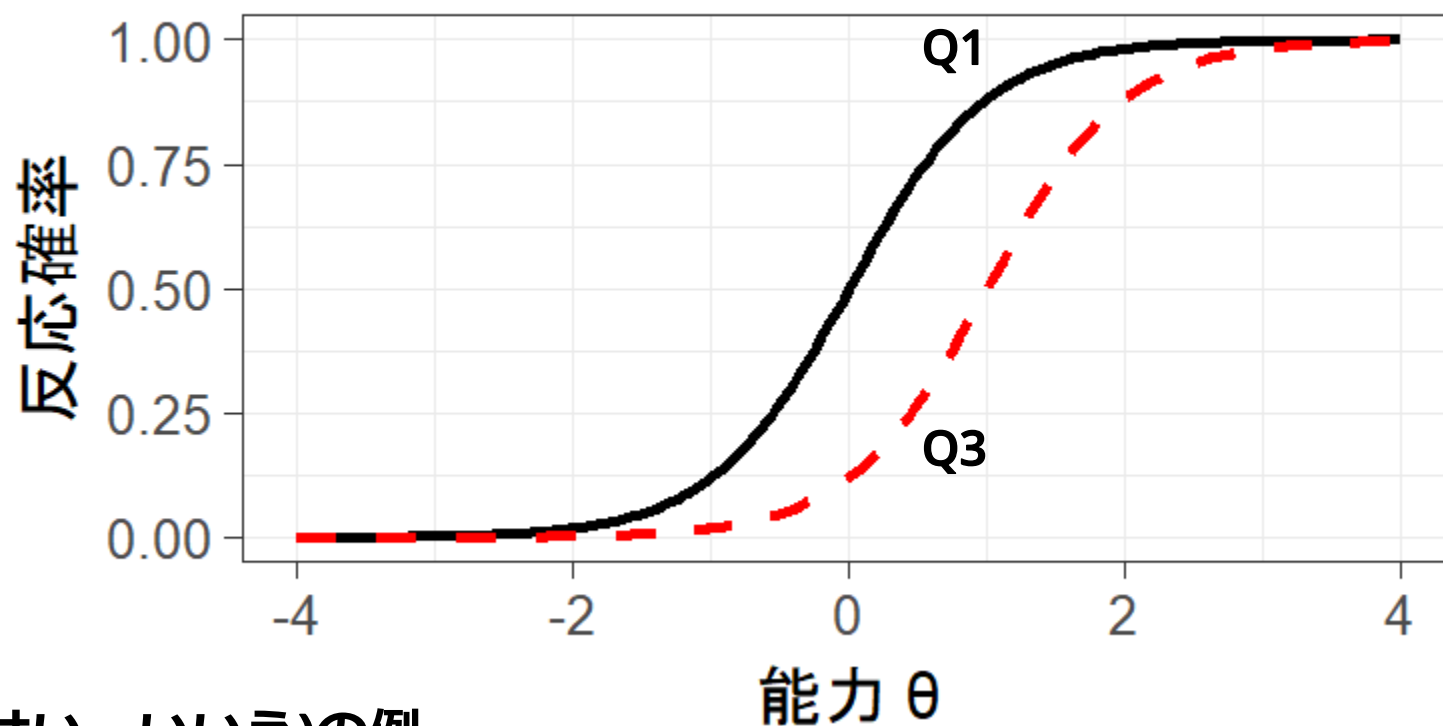
※2値(はい・いいえ)の例

2. 項目の何がわかるのか? (困難度)

ID	Q1	Q2	Q3
1	1	0	0
2	0	1	1
3	0	0	1

項目特性を推定し、
回答者の能力を推定

どれだけ 回答しやすい 項目かを検討できる(困難度)



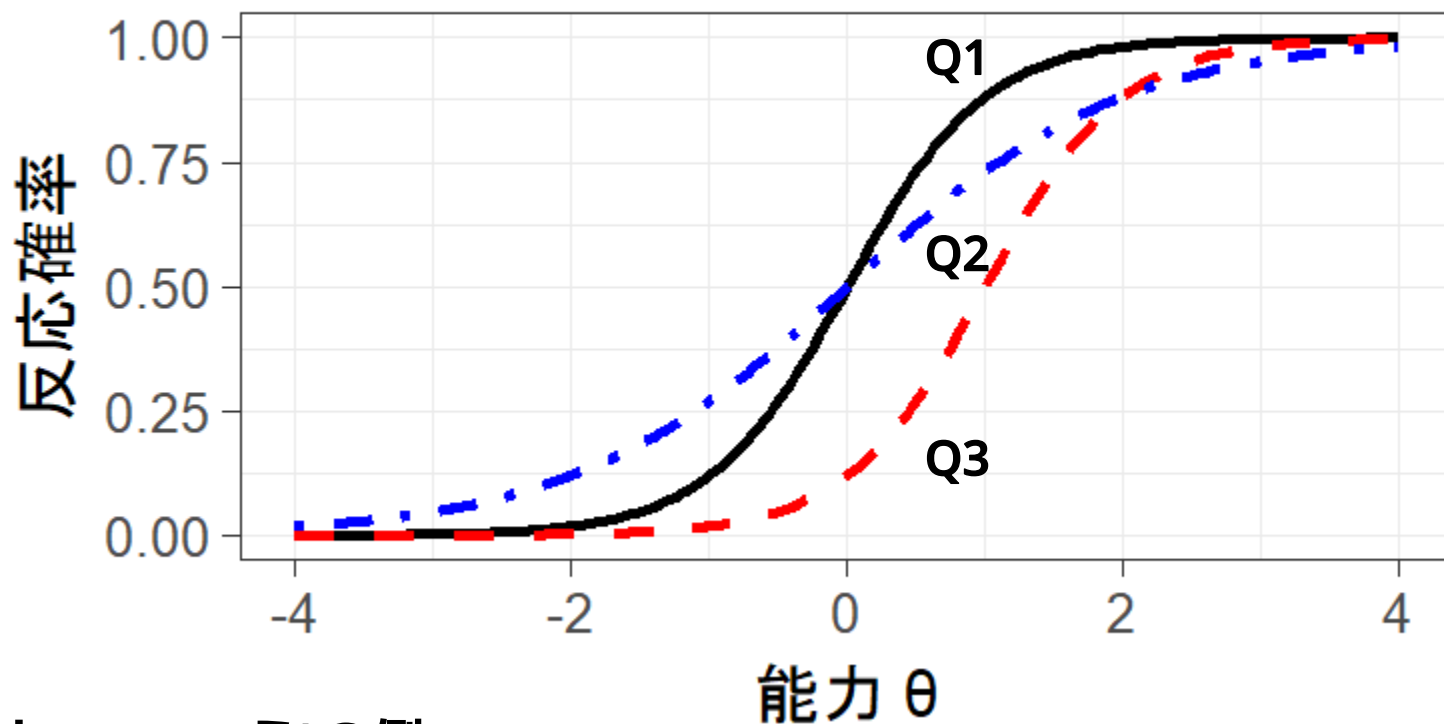
※2値(はい・いいえ)の例

2. 項目特性を推定し，能力がわかるモデル

ID	Q1	Q2	Q3
1	1	0	0
2	0	1	1
3	0	0	1

項目特性を推定し、
回答者の能力を推定

項目特性曲線(1つの線は1つの項目を表す)



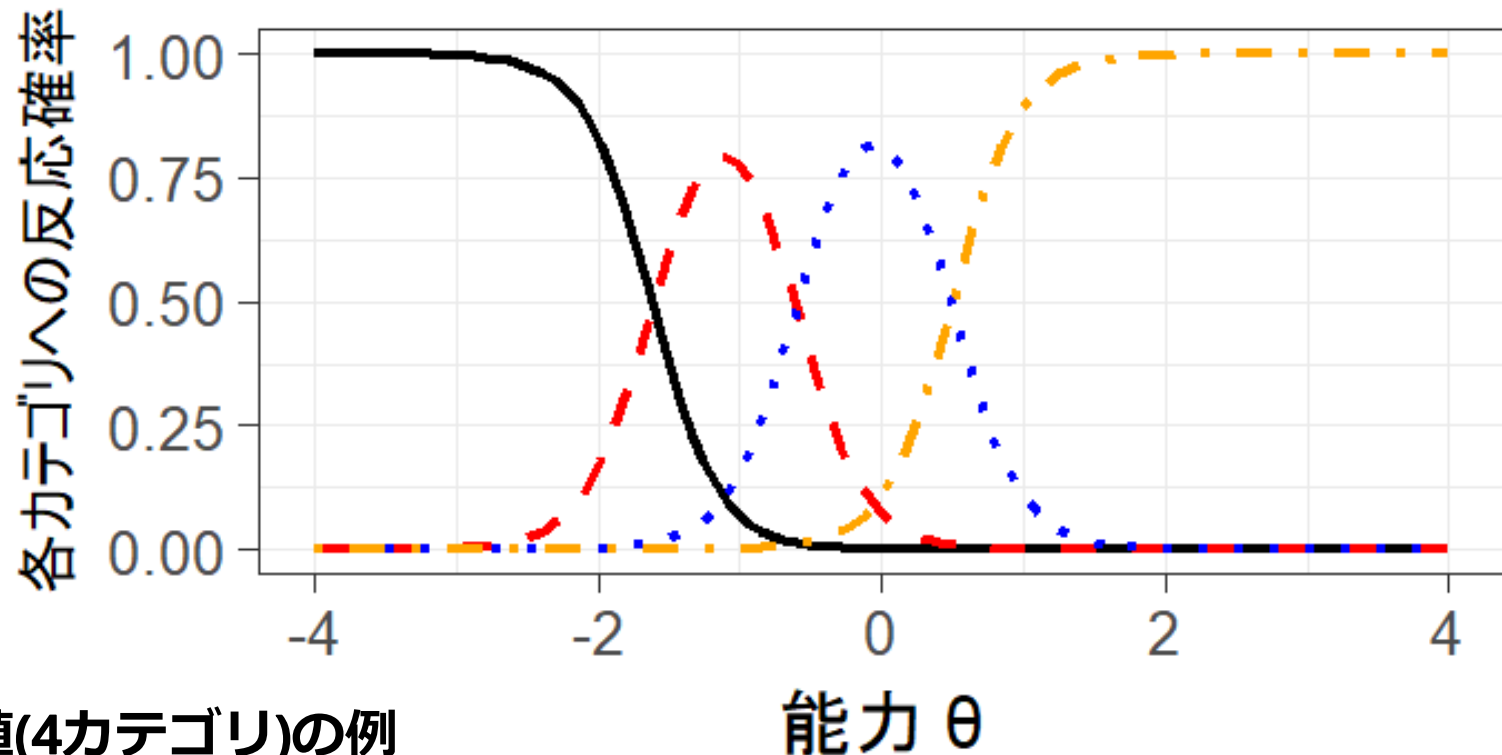
※2値(はい・いいえ)の例

2. 多値の場合は？

ID	Q1	Q2	Q3
1	1	4	3
2	4	4	3
3	1	2	2

項目特性を把握し、
回答者の能力を測定

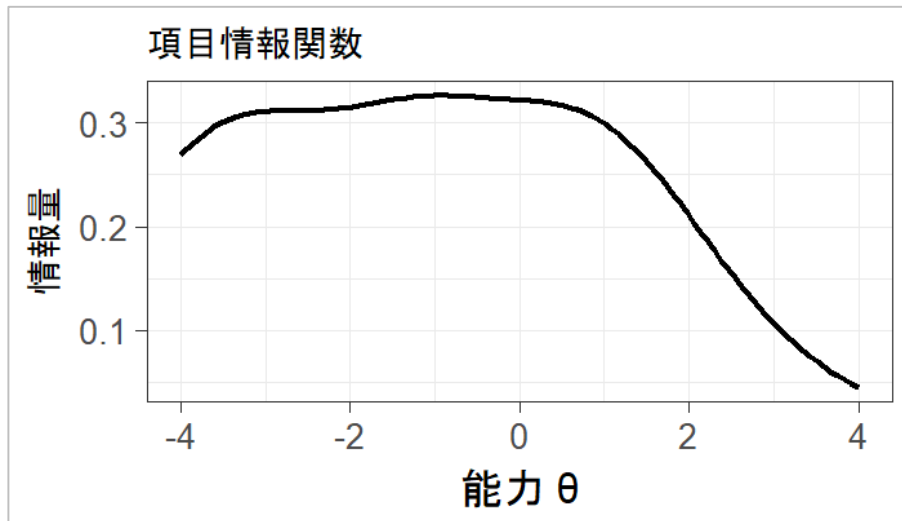
項目特性曲線(1つの線は1つのカテゴリを表す)



2. IRTのメリットは?

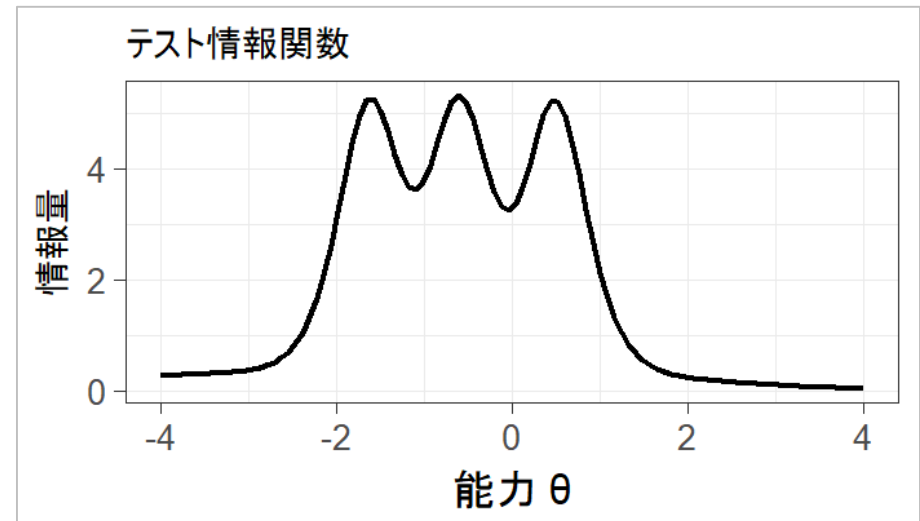
- ・項目または質問紙そのものの測定範囲を知ることができる
 - ・項目情報関数やテスト情報関数による評価が可能

項目情報関数



この**項目**が
どれくらいの能力の回答者を
測定可能かを表している

テスト情報関数



この**テスト(すべての項目)**が
どれくらいの能力の回答者を
測定可能かを表している

2. これまでのIRTを用いた臨床研究(一例)

• 短縮版の作成

- 通常版の質問紙にIRTを適用
- “良い” 項目だけを残して、短縮版を作成

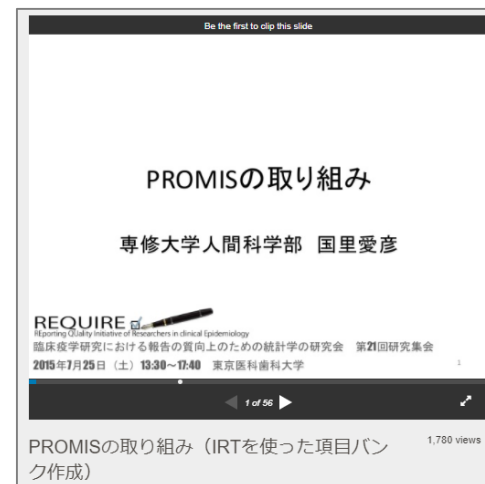
• DIF (Differential Item Functioning; 特異項目機能) の検出

- DIFとは、簡単にいえば答え方の差 (e.g. 人種・性別・臨床or非臨床)
- IRTを用いて、答え方に差が現れるか検討

• CBT (Computer Based Testing) への適用

- CBTとは、回答者の能力に適した項目が自動構成され、回答者の能力に応じた調査が行うことができる

例：PROMIS (<https://www.slideshare.net/YoshihikoKunisato/promisirt>)



今後、個人的に期待する臨床IRT研究

- **カットオフ得点からカットオフ範囲へ**
- **累積型モデルから展開型モデルへ**

カットオフ得点 からカットオフ範囲へ

- これまでの調査でのカットオフ得点
- BDI-IIを例に

0. 憂鬱ではない
1. 憂鬱である
2. いつも憂鬱から逃れることができない
3. 耐えがたいほど、憂鬱が不幸である

0. 最近それほどやせたということはない
1. 最近2kg以上やせた
2. 最近4kg以上やせた
3. 最近6kg以上やせた

⋮



1

+



1

同じ1点として解釈

＝

カットオフ
超えた? or 超えてない?

カットオフ得点 からカットオフ範囲へ

- IRTを用いる
- 得点に項目の特徴を反映させる

0. 憂鬱ではない
1. 憂鬱である
2. いつも憂鬱から逃れることができない
3. 耐えがたいほど、憂鬱が不幸である

0. 最近それほどやせたということはない
1. 最近2kg以上やせた
2. 最近4kg以上やせた
3. 最近6kg以上やせた

⋮

回答者の特性と項目の特徴に合わせて、
回答者を適切にスクリーニングする

IRT で推定した
項目の特徴
(イメージ)

$\times 1.2$



1.2

+

$\times 0.6$



0.6

||

カットオフ

超えた? or 超えてない?

項目・回答者の特性に応じて解釈

カットオフ得点 から **カットオフ範囲** へ

- ・ベイズの考え方の大きな特徴 (点ではなく範囲で考える)

実質的に等価な範囲

ROPE: Region of practical equivalence

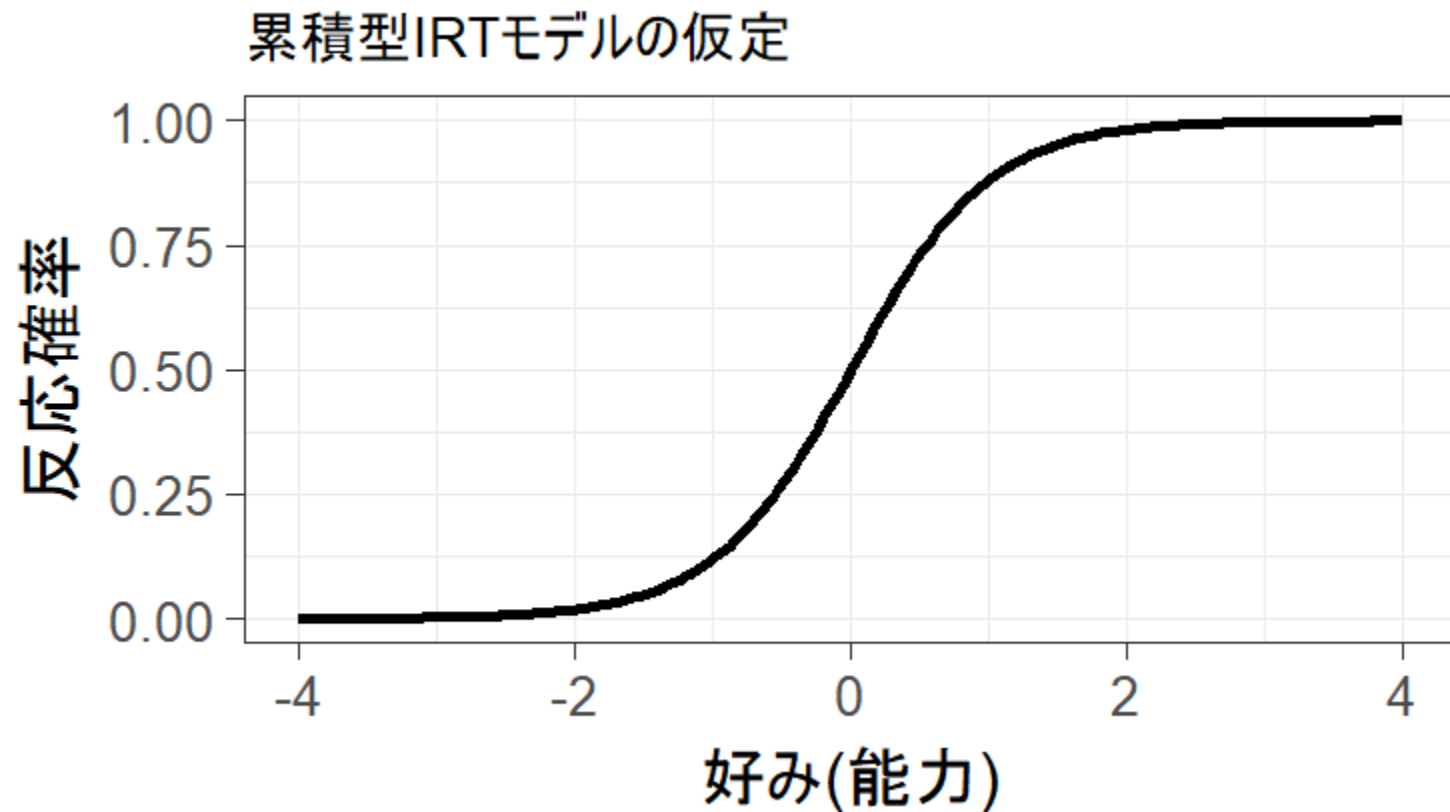
- 「等価」とみなしてよい値の範囲
- 非標準化効果量、臨床的有意性と関連する
- Ex) BDIの得点は0点-5点までの差は同じとみなす

「意味のある差」である確率はどの程度かを考えることができる

カットオフ得点からカットオフ範囲による解釈へ

累積型モデルから 展開型モデルへ

- **累積型IRTモデル** – これまで紹介してきたモデル
 - 回答者の能力は**単調増加**するという仮定
 - 例えば、**辛いものは辛くなればなるほど 好き**

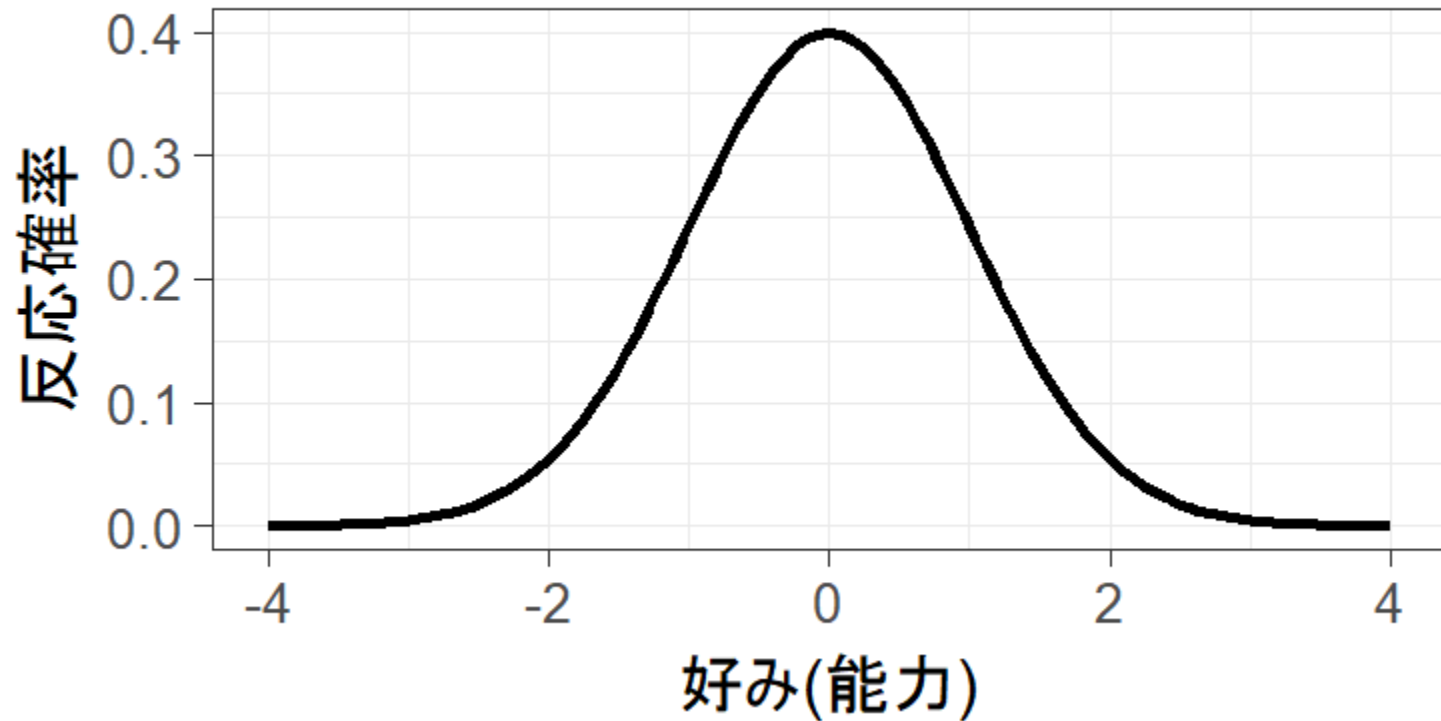


累積型モデルから 展開型モデル へ

・展開型IRTモデル

- ・回答者の能力が山型(ピーク)であるという仮定
- ・例えば、辛いものは **好き** だけど辛すぎるものは **嫌い** になる

展開型IRTモデルの仮定



好みや態度測定の心理の文脈に適している

調査における(ベイズ) 統計データ分析のまとめ

- まずは、素点で解釈する方法(古典的テスト理論)から**回答者や項目特性を考慮した方法(IRT)**へとシフト
 - 項目の善し悪しを評価することができる
 - 集団特性に依存しない回答者の解釈が可能
- 項目反応理論は頻度論的枠組みでもできるが、**ベイズ論的枠組みでやることにもメリット**あり
 - パラメータ推定が安定する
 - 知りたいことを点ではなく**範囲**で評価することができる(カットオフ等)
 - **展開型IRTモデル**で実データに即した分析をしてあげる

3. 調査研究のための ベイズ統計モデリング

回答スタイルバイアスについて

係留ビネット法 (Anchoring Vignettes method)

ベイズ多次元IRTモデル

2. これまでのIRTを用いた臨床研究(再掲)

• 短縮版の作成

- 通常版の質問紙にIRTを適用
- “良い” 項目だけを残して、短縮版を作成

• **DIF (Differential Item Functioning; 特異項目機能) の検出**

- DIFとは、簡単にいえば答え方の差 (e.g. 人種・性別・臨床or非臨床)
- IRTを用いて、答え方に差が現れるか検討

今回は、これに着目

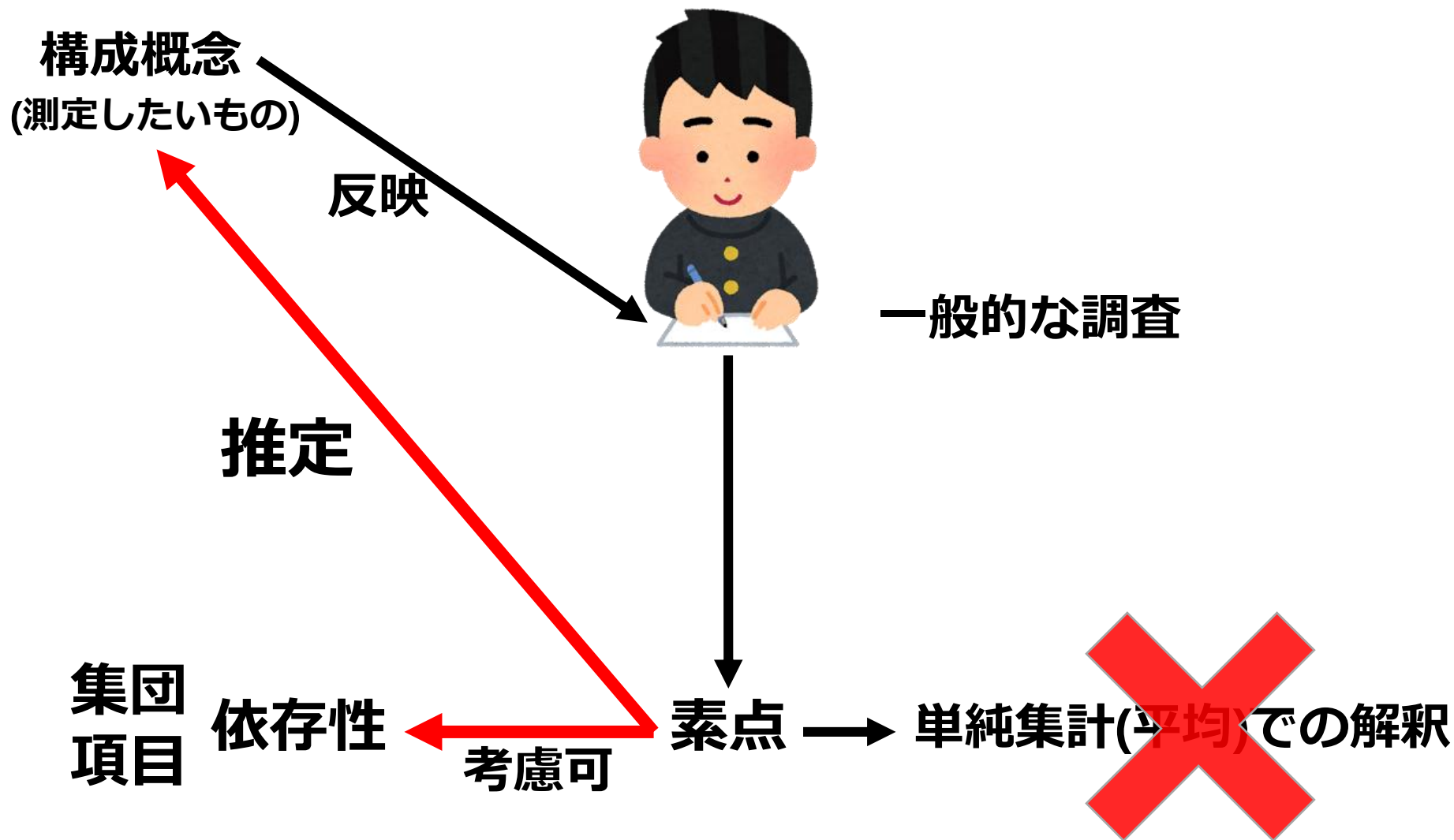
• CBT (Computer Based Testing) への適用

- CBTとは、回答者の能力に適した項目が自動構成され、回答者の能力に応じた調査が行うことができる

例: PROMIS(<https://www.slideshare.net/YoshihikoKunisato/promisirt>)

**DIF(回答バイアス)を組み込んだ
ベイズ統計モデリングを行う**

調査+IRT のイメージ (再掲)



集団・項目の特徴を考慮し、**能力**を把握する

素点ってもっといろいろ反映されているよね？

構成概念
(測定したいもの)

回答バイアス

反映



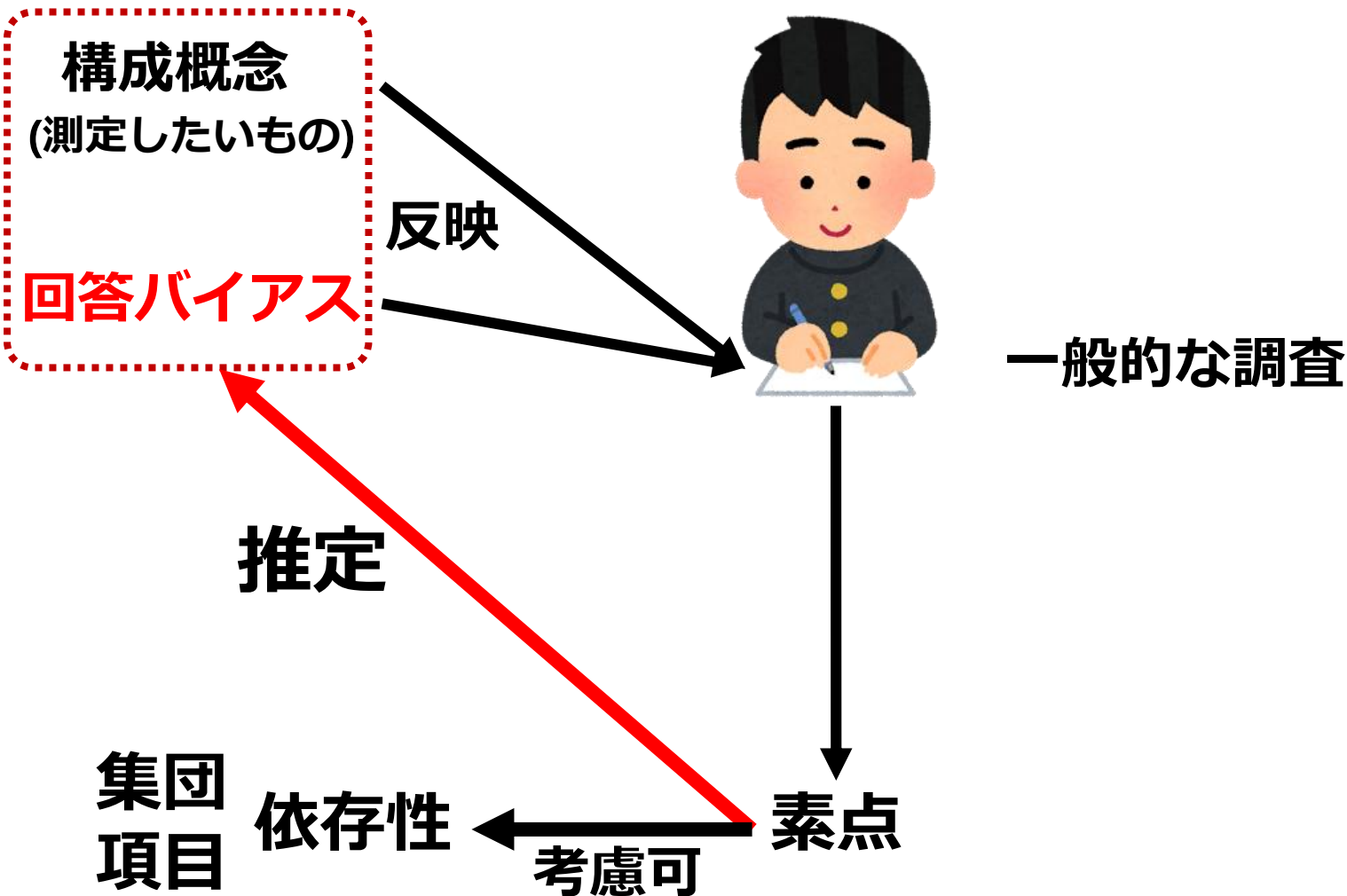
一般的な調査

集団
項目

依存性

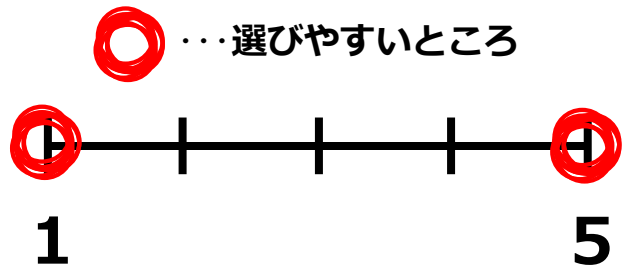
素点

ここからのお話



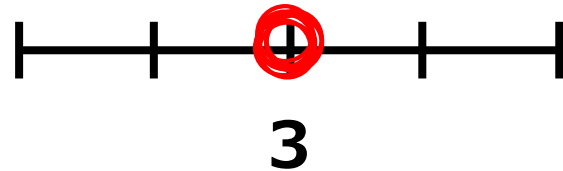
集団・項目の特徴と回答バイアス補正して
能力を把握する

3. 主な回答バイアス (5件法)



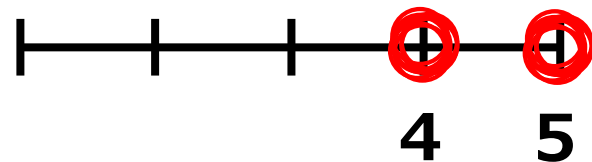
極端回答傾向

⇒ 1, 5(極端なカテゴリ)を回答しやすい



中間回答傾向

⇒ 3 (中間のカテゴリ)を回答しやすい



黙従回答傾向

⇒ 4, 5 (ポジティブなカテゴリ)を回答しやすい

当てはまらない

やや当てはまらない

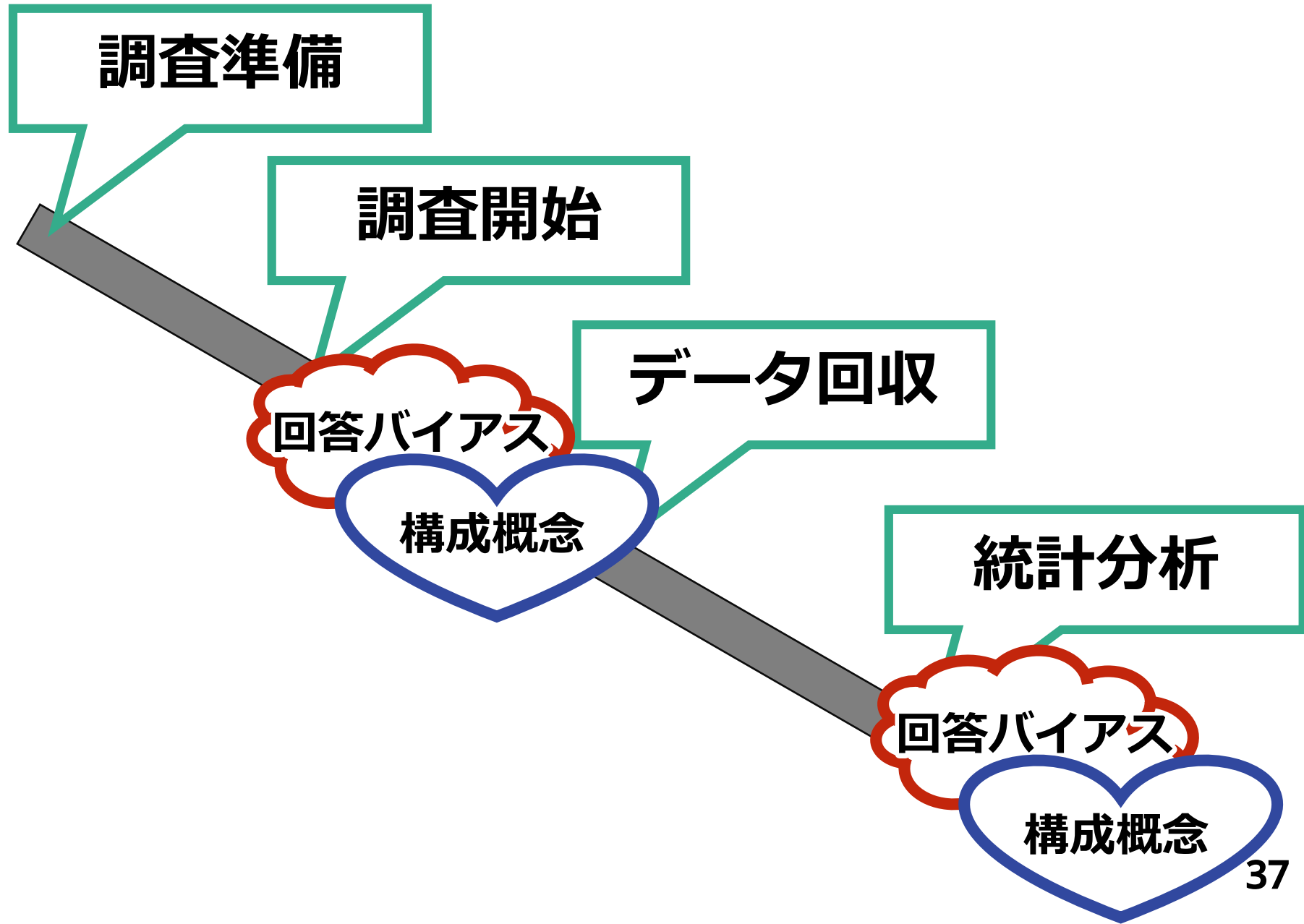
どちらでもない

やや当てはまる

とても当てはまる

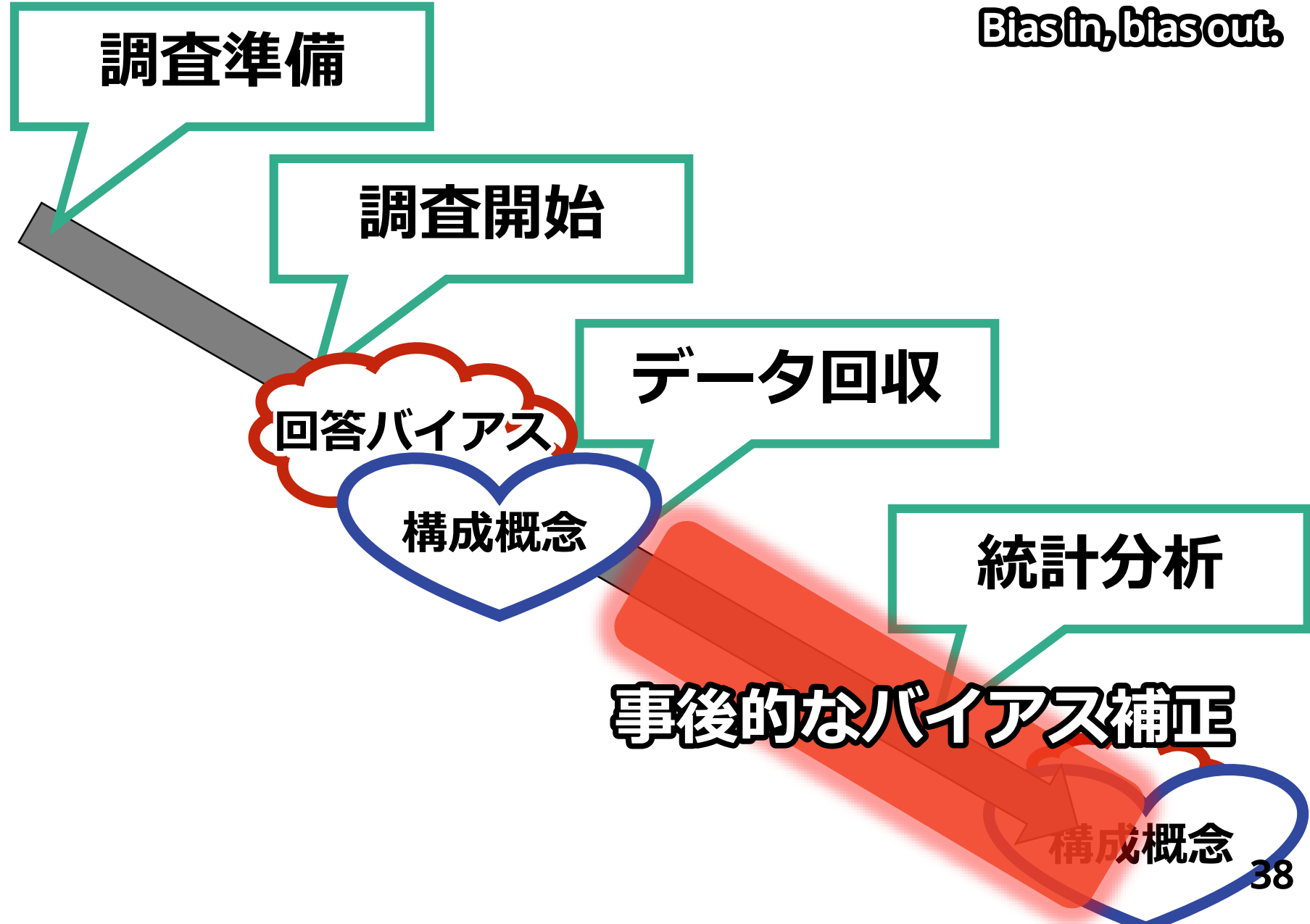
3. 一般的な調査

➡ 回答バイアスが入ったまま
結果解釈を行っている



3. これまでのDIF対処法 ➡ 事後的にどれだけバイアス補正しても混入したものを完全に除くのは難しい

Bias in, bias out.



3. どうすればよい?



回答バイアスを直接測定可能な
調査デザインにしよう

調査準備

調査開始

回答バイアスを直接
測定可能な調査デザイン

データ回収

構成概念

統計分析

事後的なバイアス補正

構成概念

我々が提案する新しい心理・臨床調査デザイン

調査準備

調査開始

けいりゅう
係留ビネット法

データ回収

統計分析

ベイズ多次元IRTモデル

係留ビネット法 (*Anchoring Vignettes method*)

回答者(あなた)の回答項目とは別に
仮想人物(A,B...)についての複数のビネットを用意し、
回答者(あなた)に読ませ、その仮想人物への回答も得る手法

ビネット項目×3

回答バイアスのみ反映

Aさんは、仕事を楽しみ、

日々の活動に支障を来させてはいない。

Bさんは、仕事に不満で、

Q-1 ときどき活動に支障を来させている。

Q-2 Cさんは、仕事ができないほど具合が悪いことがあり、

日々の活動に大きな支障を来たしている。

Q-3 あなたはCさんがどれだけ落ち込んでいると感じますか？

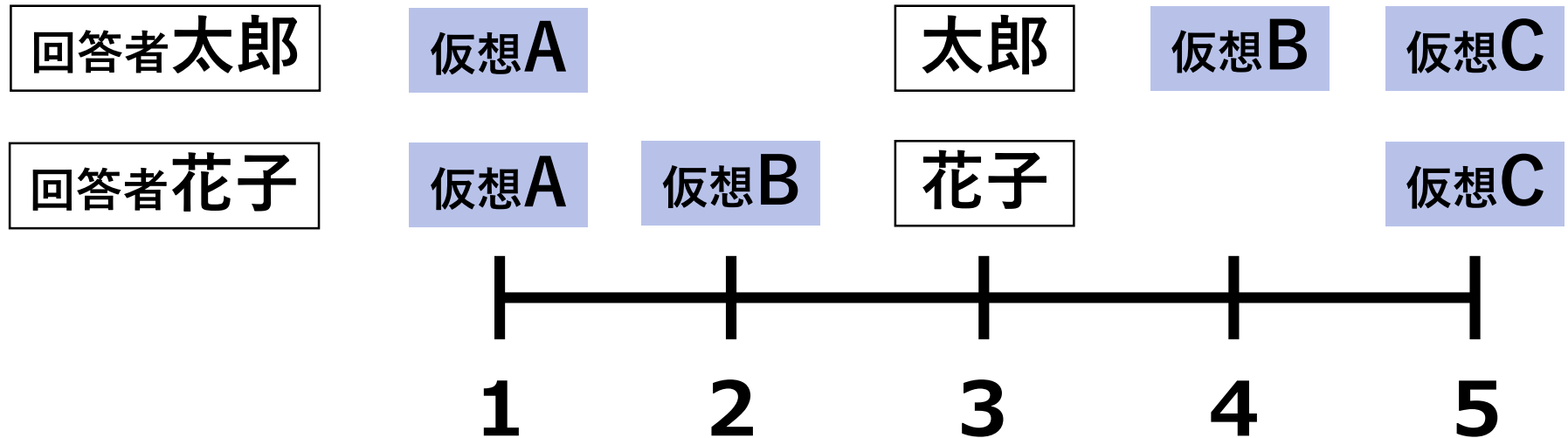
1 2 3 4 ⑤

自己評定項目 構成概念+回答バイアスを反映

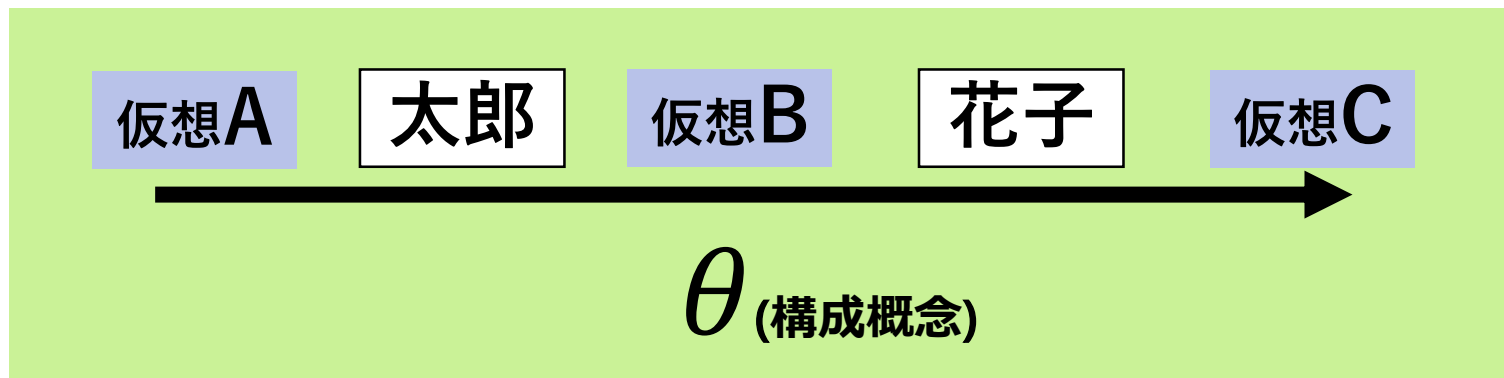
Q-4 あなたは自分がどれだけ落ち込んでいると感じますか？

1 2 3 ④ 5

係留ビネット法



多次元IRT型モデル



3. 実データ分析

- 今回使用するデータ

Select Country ▼

SHARE
Survey of Health, Ageing and Retirement in Europe
50+ in Europe

Survey of Health, Ageing and Retirement in Europe

Home

Data Access & Documentation

You are here: Data Access & Documentation - Research Data Center & Data Access

Research Data Center & Data Access

Search

search

© 2016 - CentERdata - Institute for data collection and research

- ヨーロッパ各国の健康・就労に関する大規模な調査データ
- 実際に**係留ビネット法に基づいたデータ収集**が行われている

1. Overall in the last 30 days, how much of bodily aches or pains did you have?

None

☐₁

Mild

☐₂

Moderate

☐₃

Severe

☐₄

Extreme

☐₅

3. 分析データ

• データ構造

痛み/抑うつ/睡眠/
記憶/呼吸/動作

6 因子

×

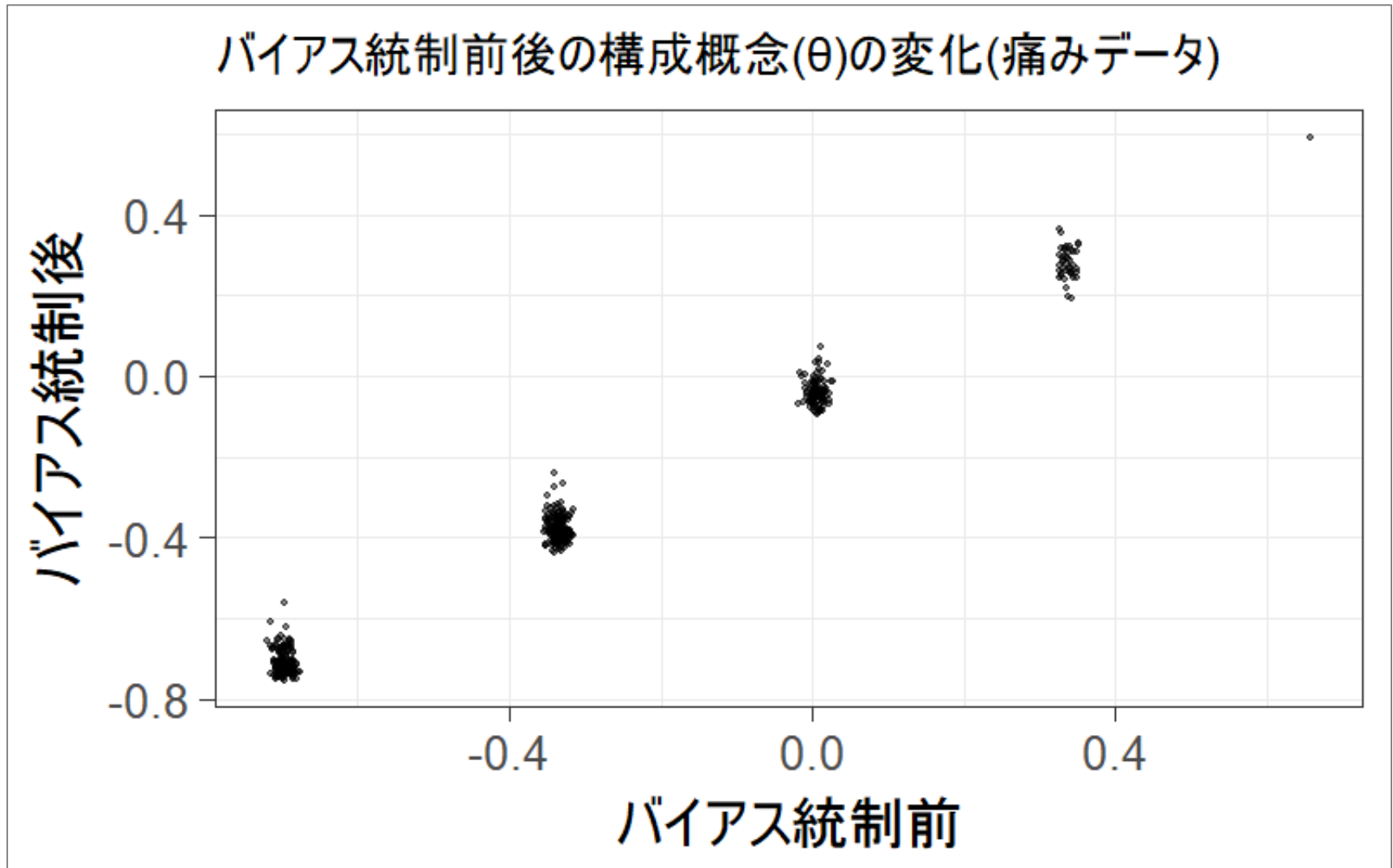
自己評価項目1項目
ビネット項目3項目

4項目

• 設定

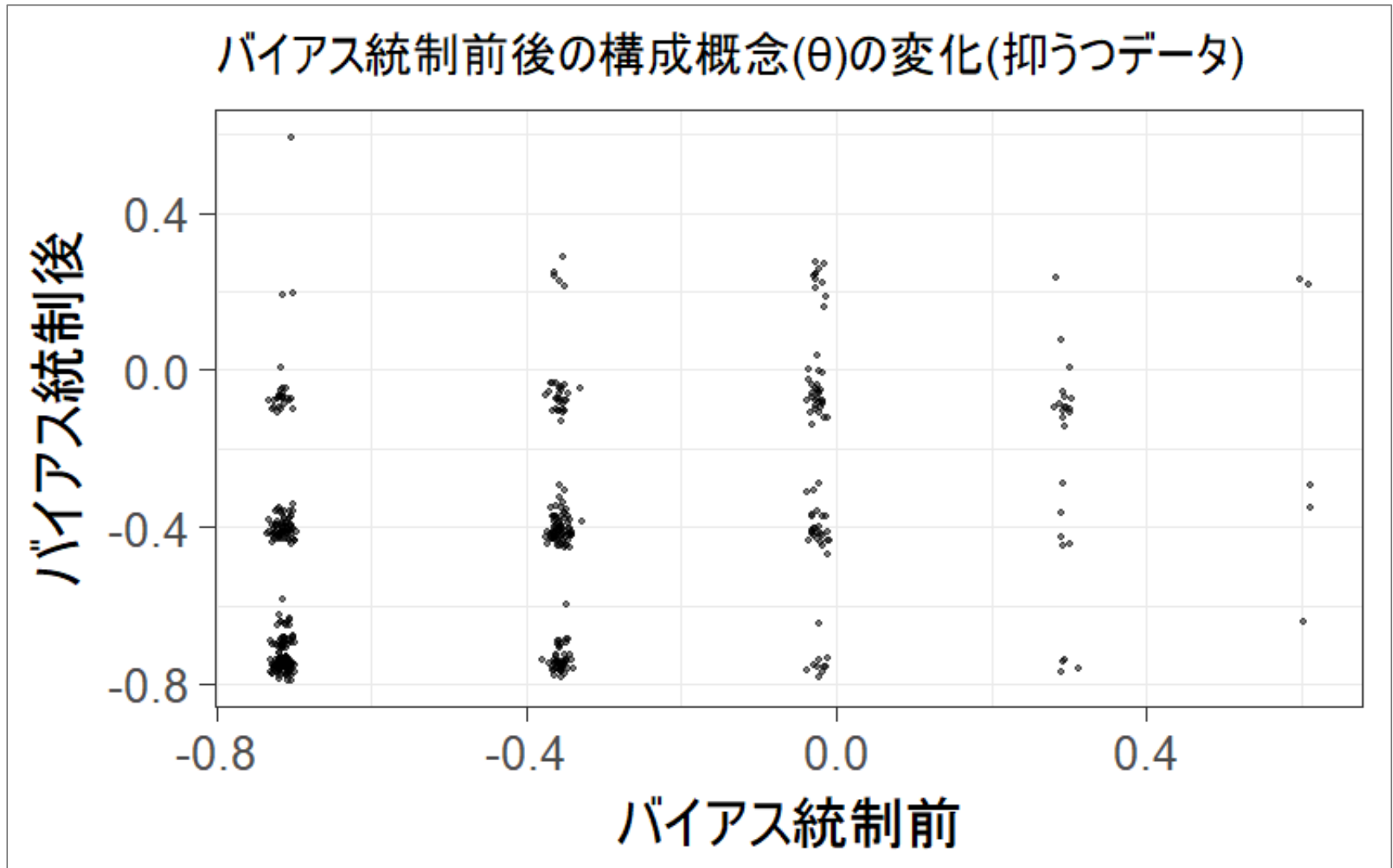
- ビネットの種類が各2種類(A or B)存在、今回はAタイプの回答者を対象
- 欠損はリストワイズ除去
- $N = 500$
- 今回は痛みと抑うつデータを使用

3. バイアス統制前後でどれだけ θ が変化するか?



痛みデータでは、回答バイアス統制しても変化はなさそう...

3. バイアス統制前後でどれだけ θ が変化するか?



抑うつデータは、回答バイアス統制することで解釈がかわる

3. 回答バイアスの推定値(抑うつ データ)

最初の10人分の回答傾向 s_r の事後平均値

各回答者の解釈

値が大きい(正)ほどそのカテゴリに回答しやすく、小さい(負)ほど回答しにくい。

0はバイアスがない

ID	各項目カテゴリにおける反応傾向バイアス						
	1	2	3	4	5		
1	0.32	0.10	0.15	-0.29	-0.27	→	1を回答しやすい
2	-0.44	0.08	0.48	-0.33	0.20	→	3を回答しやすい
3	0.01	0.47	0.16	-0.32	-0.27	→	2を回答しやすい
4	-0.08	-0.28	-0.22	0.85	-0.25	}	4を回答しやすい
5	-0.10	-0.28	-0.22	0.84	-0.26		
6	0.31	0.09	-0.24	0.09	-0.26	→	1を回答しやすい
7	-0.07	-0.29	0.51	0.11	-0.26	→	3を回答しやすい
8	-0.02	0.42	-0.25	-0.34	0.17	}	2を回答しやすい
9	-0.37	1.14	-0.20	-0.29	-0.26		
10	-0.39	-0.25	0.84	0.10	-0.27	→	3を回答しやすい

調査におけるベイズ統計モデリングのまとめ

- **DIF(回答バイアス)を取り込んだモデリングを実施**
 - 回答バイアスを取り込むために**係留ビネット法**を導入
 - 分析モデルも係留ビネット法に適したモデルに拡張
 - 結果的に回答者の構成概念と回答バイアスを分離できた
 - **痛み** データでは、回答バイアスの影響は小さかった
 - **抑うつ** データでは、回答バイアスの影響がありそう
- 今回は考慮していないが...
 - 国家別・性別等の要因をモデルに組み込んで、**より複雑なモデリングを行うことも可能**
 - 例) ある国の男性は、抑うつ質問紙を答えるときに回答バイアスの影響が大きくなってしまう

まとめ：
臨床心理学調査で
ベイズを用いるためには？

臨床心理学調査でベイズを用いるためには？

- まずは、簡単な(既存の)モデルから始めましょう
(それがベイズモデルである必要はない)
 - 例えば、素点解釈からIRTモデルを用いて分析してみる
→ **項目の特性・回答者の能力を推定してみる**
- それをベイズ論的枠組みに拡張しよう
 - 例えば、IRTモデルを**ベイズ推定**に変えてみる
→ カットオフを得点ではなく、範囲(事後分布)で考える
- これまでの流れを更に拡張してみよう
 - 例えば、取得するデータ・使用データを増やしてみる
→ データを基にモデルを構築する
→ モデル構築のポイントは**ヒストグラム・散布図を描くこと**
→ 作った複雑なモデルの推定には、**ベイズ推定が大いに活躍してくれる**
→ **そのモデルが妥当かどうかは議論を重ねる必要がある**

さいごに

- 今回は、シンポジウムということで要点を絞って広く浅く話しました。
- 海外の心理学者は積極的にベイズを用いて研究しています。その中でも多いのが、**臨床研究者と心理統計研究者がタッグ**を組んでいるケースが見受けられます。
- もし、分析したいデータはあるがどう分析していいかわからない。そんな場合は**お近くの心理統計研究者**にご相談してみてはいかがでしょうか？
- (ちなみに、私、心理統計研究者です。)

謝辞

本講演の一部は、博士後期課程2年 二瓶正登さんとの臨床－統計勉強会資料を基に作成いたしました。

連絡先: dhojou@psy.senshu-u.ac.jp

本日の資料: <https://dastatis.github.io/index.html>



回答者 r が項目 i に対して k 番目のカテゴリを選択する確率を

$$P(U_{ri} = k | \theta_r, \mathbf{s}_r) = \frac{\exp(A_{ik}\theta_r + s_{rk} + c_{ik})}{\sum_{h=1}^K \exp(A_{ih}\theta_r + s_{rh} + c_{ih})}$$

where $\sum_h s_{rh} = 0$ and $\sum_h c_{ih} = 0$.

$$A_{ik}\theta_r + s_{rk} + c_{ik}$$

カテゴリ間隔 × 構成概念 + 回答バイアス + 切片

回答バイアスのみ

ビネット項目

$$s_{rk} + c_{ik}$$

構成概念+回答バイアス

自己評定項目

$$A_{ik}\theta_r + s_{rk} + c_{ik}$$