比較型項目における
項目フォーマットと情報量の関係について

分寺 杏介[1]　岡田 謙介[2]

[1]神戸大学　経営学研究科
[2]東京大学　教育学研究科

[1]bunji@bear.kobe-u.ac.jp

Outline

1 Introduction

多肢選択のフォーマットの話

2 方法

分析モデル (Thurstonian IRT)

項目・テスト情報量

3 結果

テスト情報量の比較

単位時間あたりの比較

回答しやすさの評価

心理学でよく使われる質問の形式

リッカート尺度

一つ一つの項目について「程度」を答える

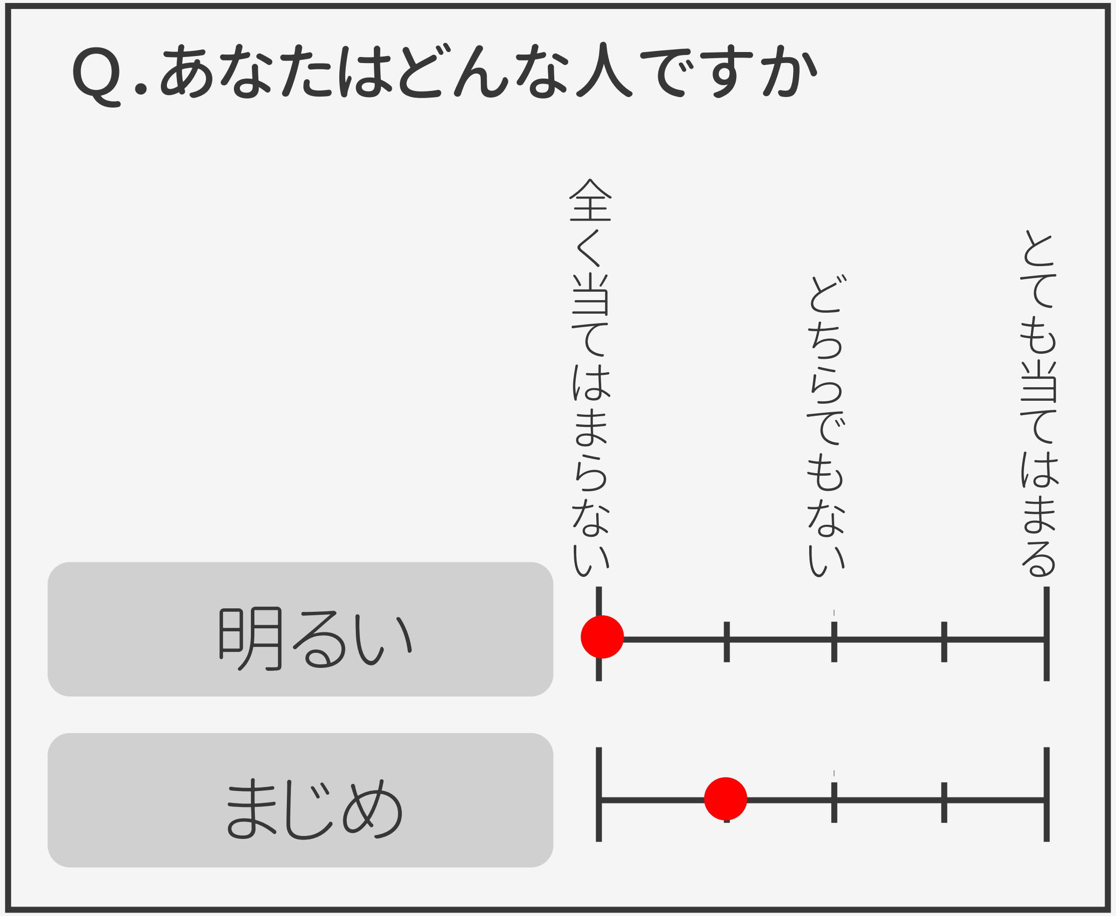
e.g., 自分に当てはまる程度，賛成する程度

系統的バイアスの影響を受けやすい

中心化傾向…5段階なら3を選びやすい

黙従傾向（寛大化傾向）…内容とは無関係に「あてはまる」を選びやすい

フェイキング…自分がよく見えるように意図的に回答を変える　など



Single-Stimulus (SS); Likert scale

バイアスが発生しにくい質問方法

Q.どっちがより当てはまる?

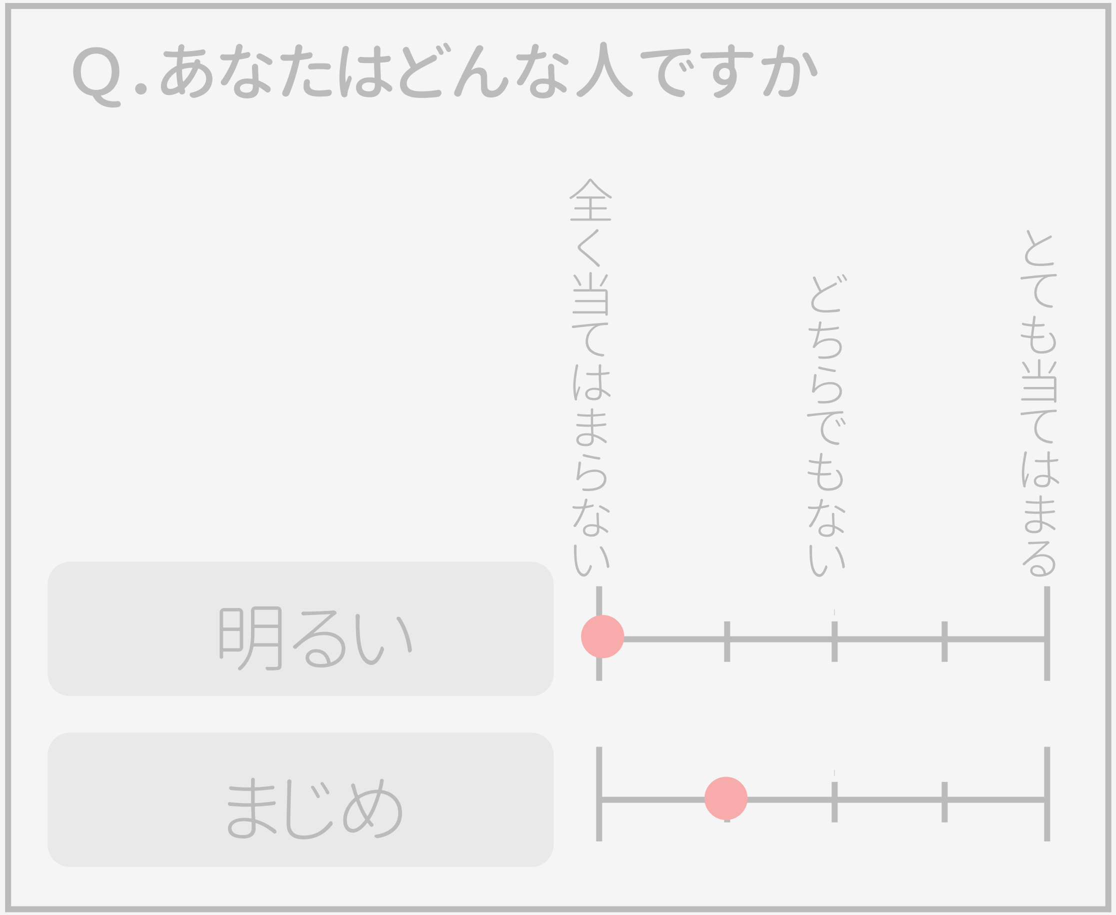
明るい　まじめ

▌(多肢)強制選択式(Forced-Choice; FC)

複数の選択肢の中から回答する形式

- 最も当てはまる／らないもの
- 当てはまる順にランキング

▶ 中心化傾向などは起こり得ない
フェイキングもかなり抑えられる

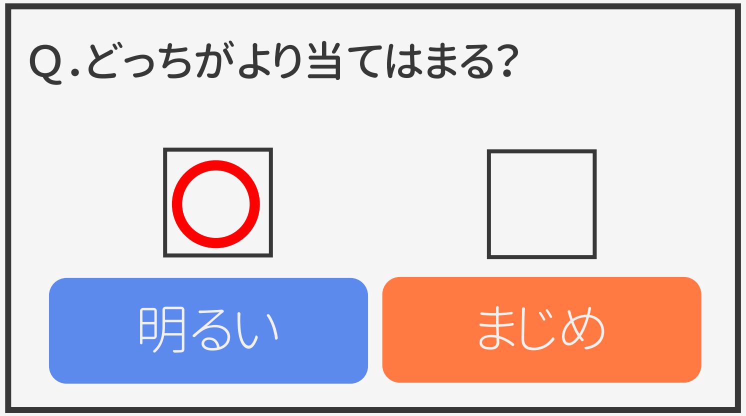


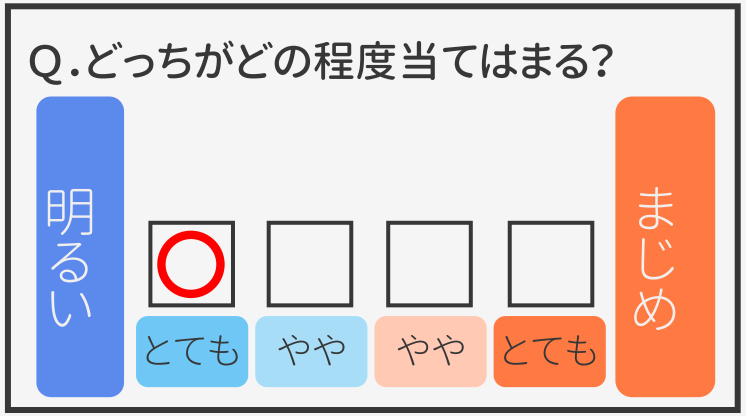
Single-Stimulus (SS); Likert scale

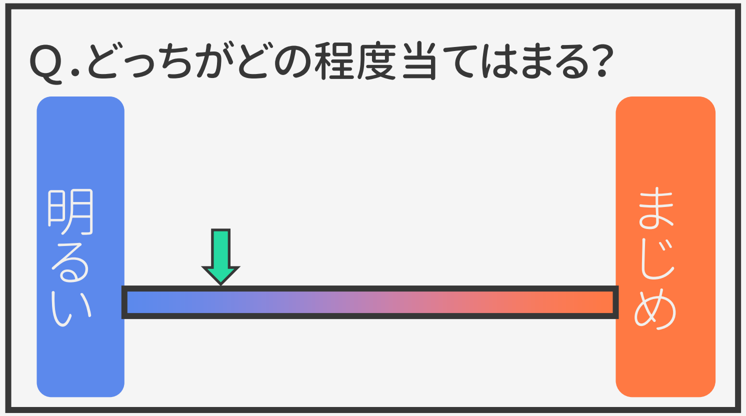
FCのほうが妥当性が高いという先行研究あり (e.g., Christiansen et al., 2005; Salgado & Táuriz, 2014)

二肢強制選択の種類

いくつかのフォーマットが考えられる







二択

答えやすい?

多肢選択

情報が多い?

スライダースケール

情報がもっと多い?

情報量

- 測定の目的は回答者の特性値を精度良く推定すること

情報量…特性値の推定に対してどの程度の情報を与えるか

- その観点で見ると…

Q.どっちがより当てはまる?

明るい　まじめ

カテゴリ数が多いほど
情報量は多い

Q.どっちがどの程度当てはまる?

明るい　とても　やや　やや　とても　まじめ

明るい >>>>>> まじめ

回答からは区別できない

明るい > まじめ

明るい >>>>>> まじめ

ある程度区別できる!

明るい > まじめ

カテゴリ数の影響

- **カテゴリ数が多いと回答が難しくなる?**

  認知負荷が高くなるかもしれない

  ▶ 回答時間が長くなるとしたら，単位時間あたり情報量はどうなる?

  リッカート尺度はVAS（スライダー）より回答しやすい(Dourado et al., 2021; van Laerhoven et al., 2004)

  ▶ 比較の場合も同じことが言えるか?

- **カテゴリ数が多いと正しく回答できないかも?**

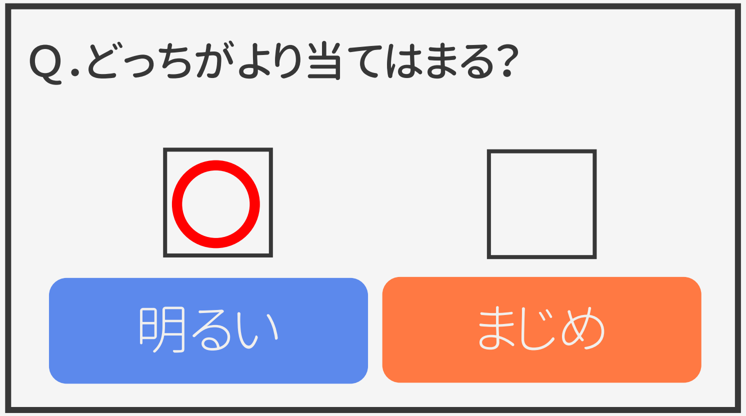
  ノイズの影響で識別力が低下するならば情報量も下がる

  リッカート尺度の場合カテゴリ数が多いほうが因子負荷は高くなる (Xu & Leung, 2018)

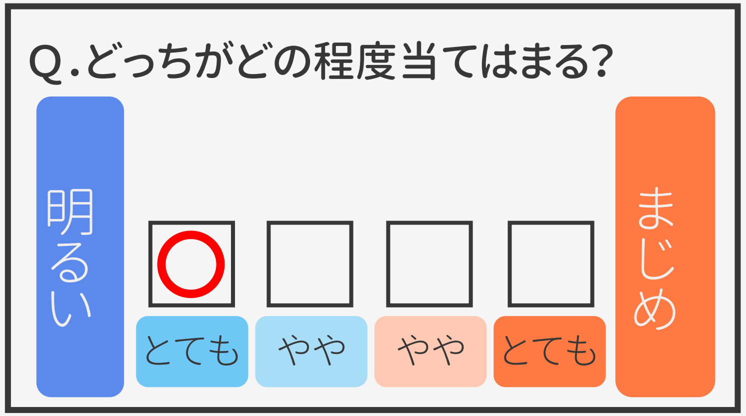
  ▶ 比較の場合も同じことが言えるか?

ここまでのまとめ

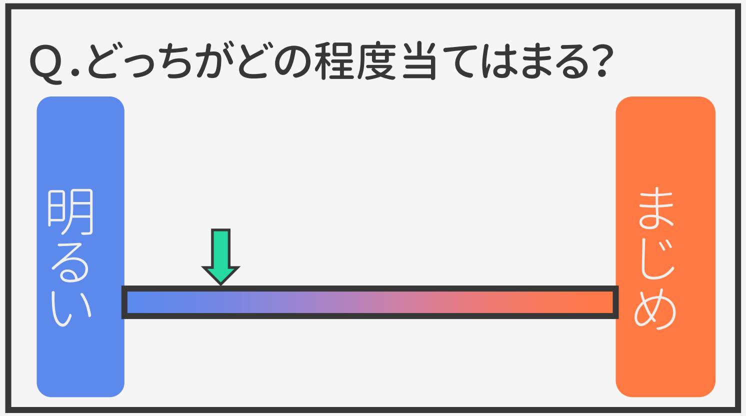
強制選択のフォーマット



二択



多肢選択



スライダースケール

理論的には

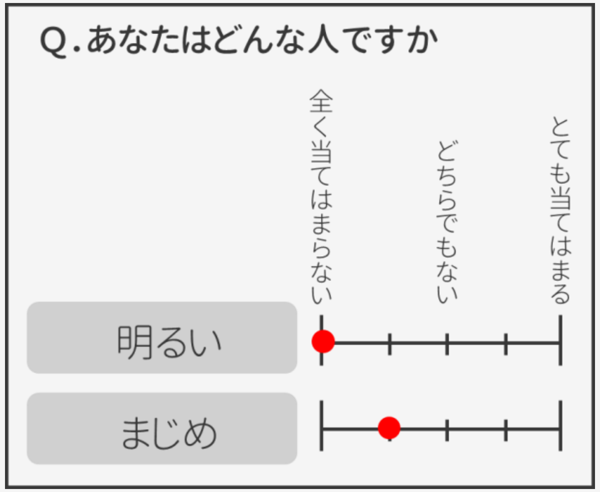
情報量	少ない	→	多い
答えやすさ	答えやすい?	→	答えにくい?
回答時間	短い?	→	長い?

研究の目的

異なるフォーマットを比較する

二択・多肢選択・スライダーにリッカートを加えた4種類

- （項目・テスト）情報量
- 単位時間あたり情報量
- （主観的な）回答しやすさ



Outline

1 Introduction

- 多肢選択のフォーマットの話

2 モデル

- 分析モデル (Thurstonian IRT)
- 項目・テスト情報量

3 結果

- テスト情報量の比較
- 単位時間あたりの比較
- 回答しやすさの評価

多肢強制選択回答を分析するモデル

Thurstonian IRTモデル (Brown and Maydeu-Olivares, 2011)

二択

Q.どっちがより当てはまる?

0	1
明るい	まじめ
$j^{(l)}$: 外向性	$j^{(m)}$: 誠実性

- 2つの文 $(j^{(l)}, j^{(m)})$はそれぞれ異なる因子 $\left(d_j^{(l)}, d_j^{(m)}\right)$ を測定する

$$X_{ij}^{(l\_m)^*} = u_{ij}^{(m)} - u_{ij}^{(l)} \qquad X_{ij}^{(l\_m)} = 1 \ \text{ if } X_{ij}^{(l\_m)^*} \geq 0$$

- 一方の潜在変数 $u_{ij}^{(m)}$ について見ると

$$u_{ij}^{(m)} = \mu_j^{(m)} + \beta_j^{(m)} \eta_{id_j^{(m)}} + \varepsilon_{ij}^{(m)} \qquad \varepsilon_{ij}^{(m)} \sim N\left(0, {\Psi_j^{(m)}}^2\right)$$

- $(j^{(l)}, j^{(m)})$ 間の比較において選択肢 $j^{(m)}$ が選ばれる確率は

正規累積モデル

$$P\left(X_{ij}^{(l\_m)} = 1 \middle| \boldsymbol{\eta}_i\right) = \Phi\left[\left(\mu_j^{(m)} - \mu_j^{(l)}\right) + \beta_j^{(m)} \eta_{id_j^{(m)}} - \beta_j^{(l)} \eta_{id_j^{(l)}}\right]$$

$\Phi(\cdot)$ ▶ 標準正規分布の累積分布関数

または

ロジスティックモデル

$$P\left(X_{ij}^{(l\_m)} = 1 \middle| \boldsymbol{\eta}_i\right) = \left(1 - \exp\left[\left(\mu_j^{(m)} - \mu_j^{(l)}\right) + \beta_j^{(m)} \eta_{id_j^{(m)}} - \beta_j^{(l)} \eta_{id_j^{(l)}}\right]\right)^{-1}$$

μ: 選択肢の選好の平均

β : 因子負荷

η : 因子得点(特性値)

TIRTのイメージ

正規累積モデル

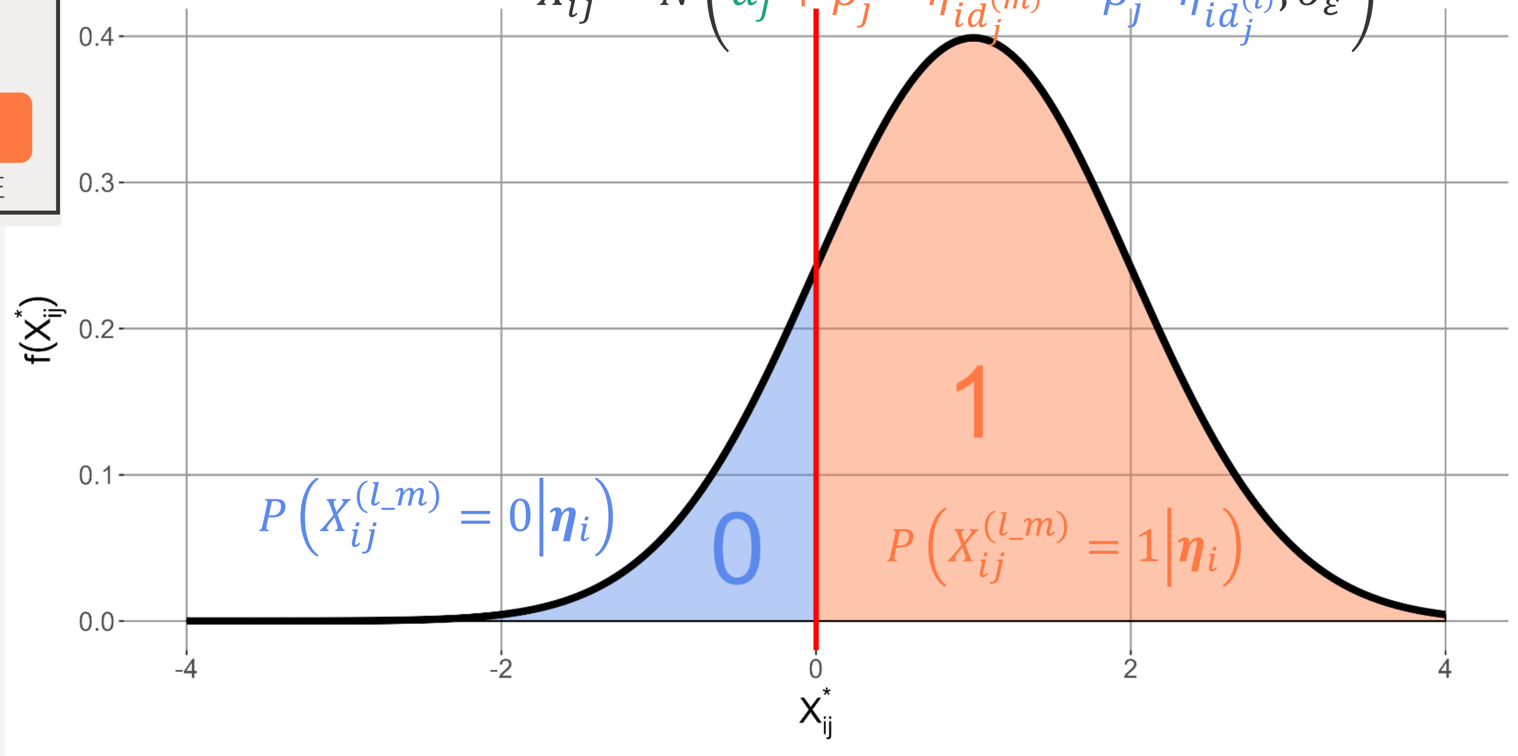
$$P\left(X_{ij}^{(l\_m)}=1\middle|\boldsymbol{\eta}_i\right)=\Phi\left[\left(\mu_j^{(m)}-\mu_j^{(l)}\right)+\beta_j^{(m)}\eta_{id_j^{(m)}}-\beta_j^{(l)}\eta_{id_j^{(l)}}\right]$$

α_j

Q.どっちがより当てはまる?

0	1
明るい	まじめ
$j^{(l)}$: 外向性	$j^{(m)}$: 誠実性

$$X_{ij}^{*}\sim N\left(\alpha_j+\beta_j^{(m)}\eta_{id_j^{(m)}}-\beta_j^{(l)}\eta_{id_j^{(l)}},\sigma_\varepsilon^2\right)$$



$\mu_j^{(l)}+\beta_j^{(l)}\eta_{id_j^{(l)}}$
が大きい

文 $j^{(l)}$ が
選ばれやすい

$\mu_j^{(m)}+\beta_j^{(m)}\eta_{id_j^{(m)}}$
が大きい

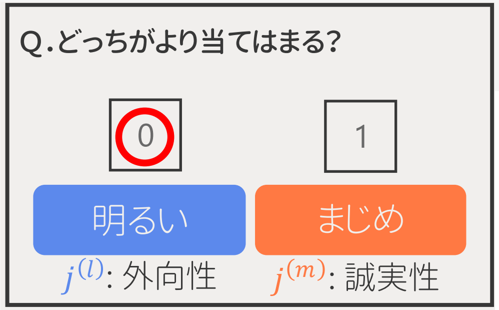
文 $j^{(m)}$ が
選ばれやすい

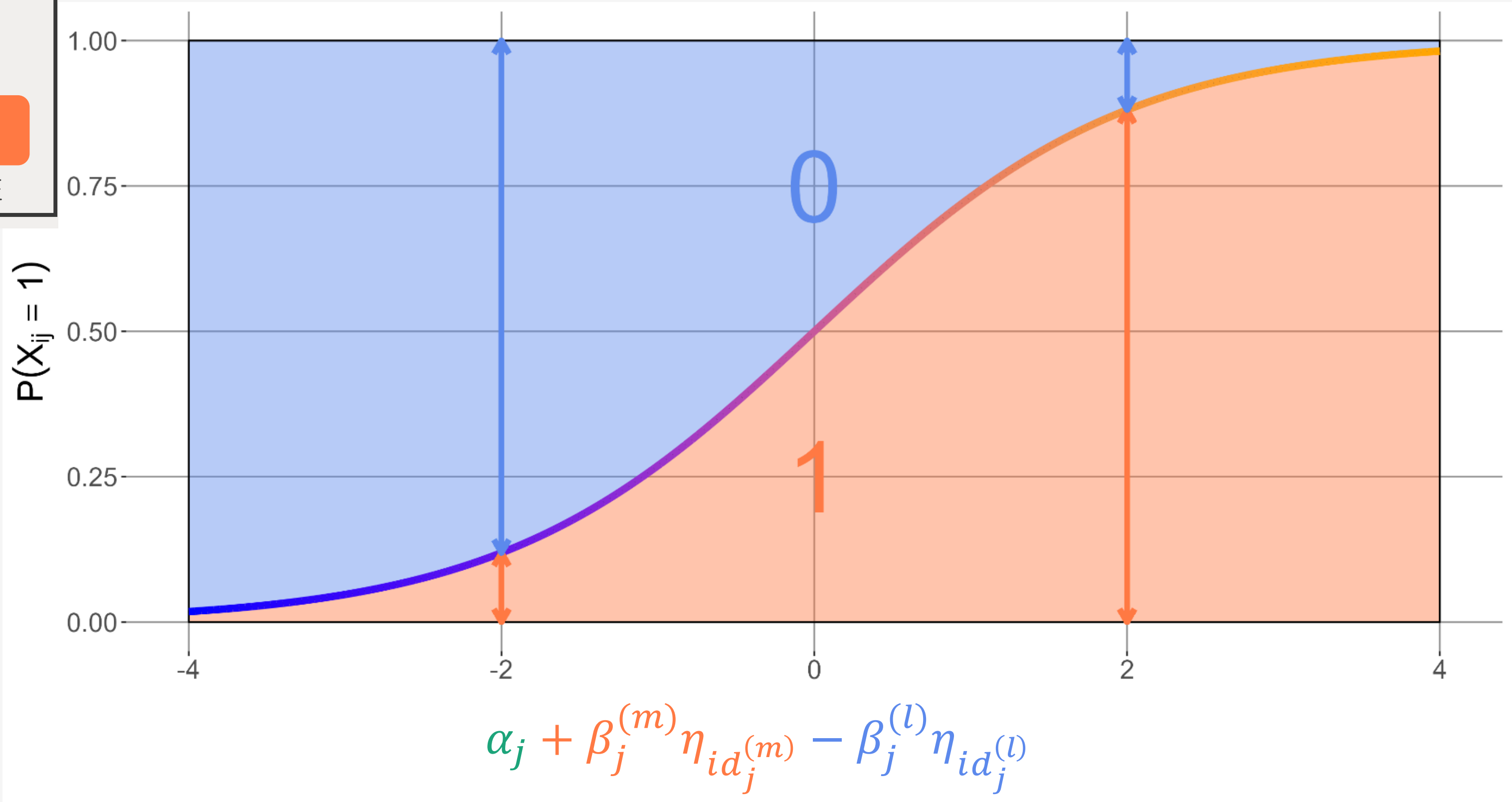
$$X_{ij}^{*}=\alpha_j+\beta_j^{(m)}\eta_{id_j^{(m)}}-\beta_j^{(l)}\eta_{id_j^{(l)}}+\left(\varepsilon_{ij}^{(l)}-\varepsilon_{ij}^{(m)}\right)$$

TIRTのイメージ

正規累積モデル

$$P\left(X_{ij}^{(l\_m)}=1\middle|\boldsymbol{\eta}_i\right)=\Phi\left[\underbrace{\left(\mu_j^{(m)}-\mu_j^{(l)}\right)}_{\alpha_j}+\beta_j^{(m)}\eta_{id_j^{(m)}}-\beta_j^{(l)}\eta_{id_j^{(l)}}\right]$$





$\mu_j^{(l)}+\beta_j^{(l)}\eta_{id_j^{(l)}}$ が大きい

文 $j^{(l)}$ が選ばれやすい

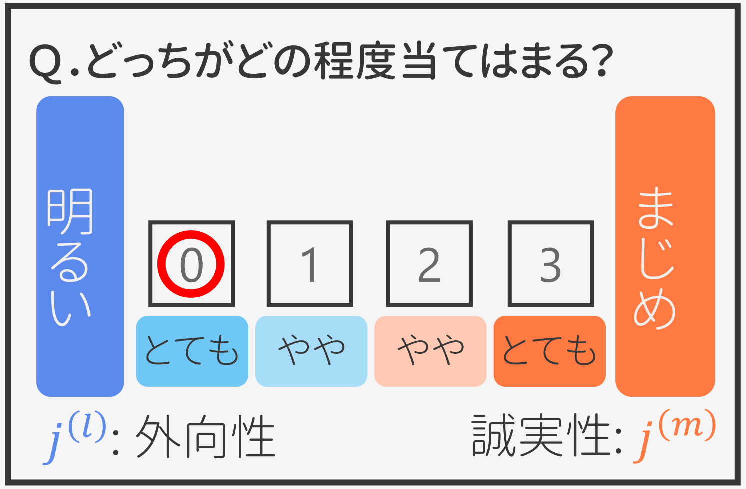
$\mu_j^{(m)}+\beta_j^{(m)}\eta_{id_j^{(m)}}$ が大きい

文 $j^{(m)}$ が選ばれやすい

多肢強制選択回答を分析するモデル

TIRTの拡張 (Brown and Maydeu-Olivares, 2018)

多肢選択



- 2つの文 $(j^{(l)}, j^{(m)})$はそれぞれ異なる因子$\left(d_j^{(l)}, d_j^{(m)}\right)$を測定する

$$X_{ij}^{(l\_m)^*} = u_{ij}^{(l)} - u_{ij}^{(m)} \qquad X_{ij}^{(l\_m)} = \begin{cases} C-1 & \text{if} \quad X_{ij}^{(lm)^*} \geq \tau_{C-1} \\ C-2 & \text{if} \quad \tau_{C-1} \geq X_{ij}^{(lm)^*} \geq \tau_2 \\ \cdots & \\ 1 & \text{if} \quad \tau_2 \geq X_{ij}^{(lm)^*} \geq \tau_1 \\ 0 & \text{if} \quad X_{ij}^{(lm)^*} \leq \tau_1 \end{cases}$$

- あとはTIRTと同じ

- $(j^{(l)}, j^{(m)})$ 間の比較において選択肢 c が選ばれる確率は

正規累積モデル

$$P\left(X_{ij}^{(l\_m)} \geq c \middle| \boldsymbol{\eta}_i\right) = \Phi\left[\alpha_j + \beta_j^{(m)} \eta_{id_j^{(m)}} - \beta_j^{(l)} \eta_{id_j^{(l)}} - \tau_c\right]$$

$$P\left(X_{ij}^{(l\_m)} = c \middle| \boldsymbol{\eta}_i\right) = P\left(X_{ij}^{(l\_m)} \geq c \middle| \boldsymbol{\eta}_i\right) - P\left(X_{ij}^{(l\_m)} \geq c+1 \middle| \boldsymbol{\eta}_i\right)$$

α: 選好の平均の差

β : 因子負荷

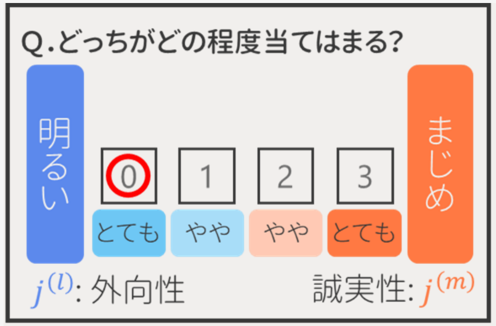
η : 因子得点(特性値)

TIRT多肢選択のイメージ

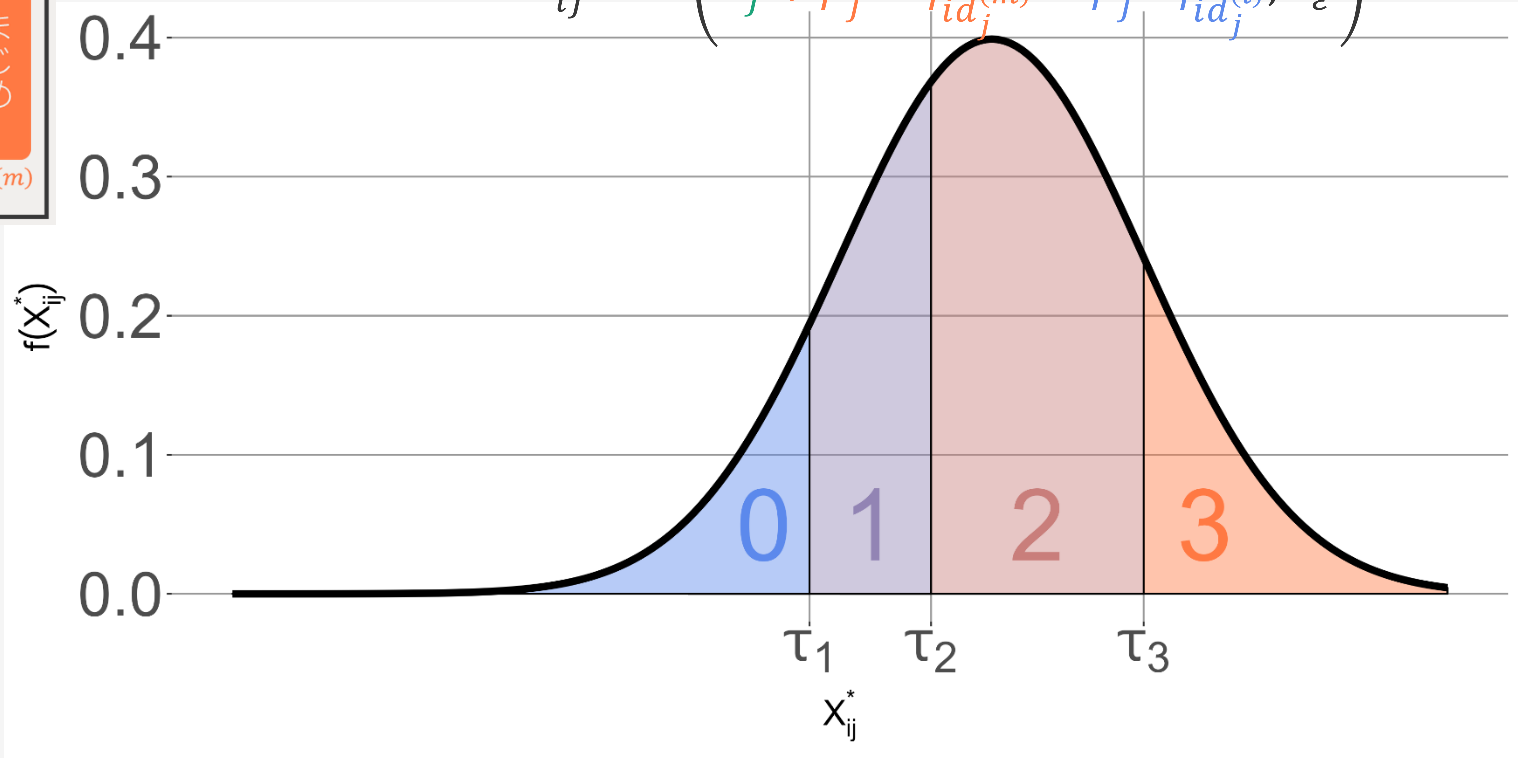
正規累積モデル

$$P\left(X_{ij}^{(l\_m)} \geq c \middle| \boldsymbol{\eta}_i\right) = \Phi\left[\alpha_j + \beta_j^{(m)}\eta_{id_j^{(m)}} - \beta_j^{(l)}\eta_{id_j^{(l)}} - \tau_c\right]$$

$$P\left(X_{ij}^{(l\_m)} = c \middle| \boldsymbol{\eta}_i\right) = P\left(X_{ij}^{(l\_m)} \geq c \middle| \boldsymbol{\eta}_i\right) - P\left(X_{ij}^{(l\_m)} \geq c+1 \middle| \boldsymbol{\eta}_i\right)$$



$$X_{ij}^{*} \sim N\left(\alpha_j + \beta_j^{(m)}\eta_{id_j^{(m)}} - \beta_j^{(l)}\eta_{id_j^{(l)}}, \sigma_\varepsilon^2\right)$$



$\mu_j^{(l)} + \beta_j^{(l)}\eta_{id_j^{(l)}}$
が大きい

文 $j^{(l)}$ が強く
選ばれやすい

$\mu_j^{(m)} + \beta_j^{(m)}\eta_{id_j^{(m)}}$
が大きい

文 $j^{(m)}$ が強く
選ばれやすい

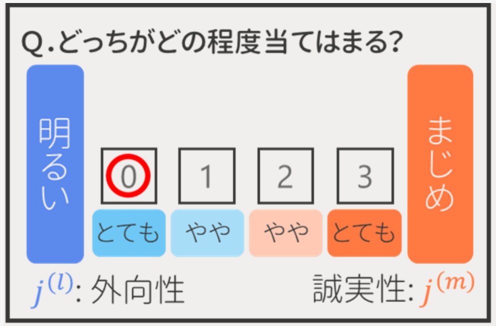
$$X_{ij}^{*} = \alpha_j + \beta_j^{(m)}\eta_{id_j^{(m)}} - \beta_j^{(l)}\eta_{id_j^{(l)}} + \left(\varepsilon_{ij}^{(l)} - \varepsilon_{ij}^{(m)}\right)$$

TIRT多肢選択のイメージ

正規累積
モデル

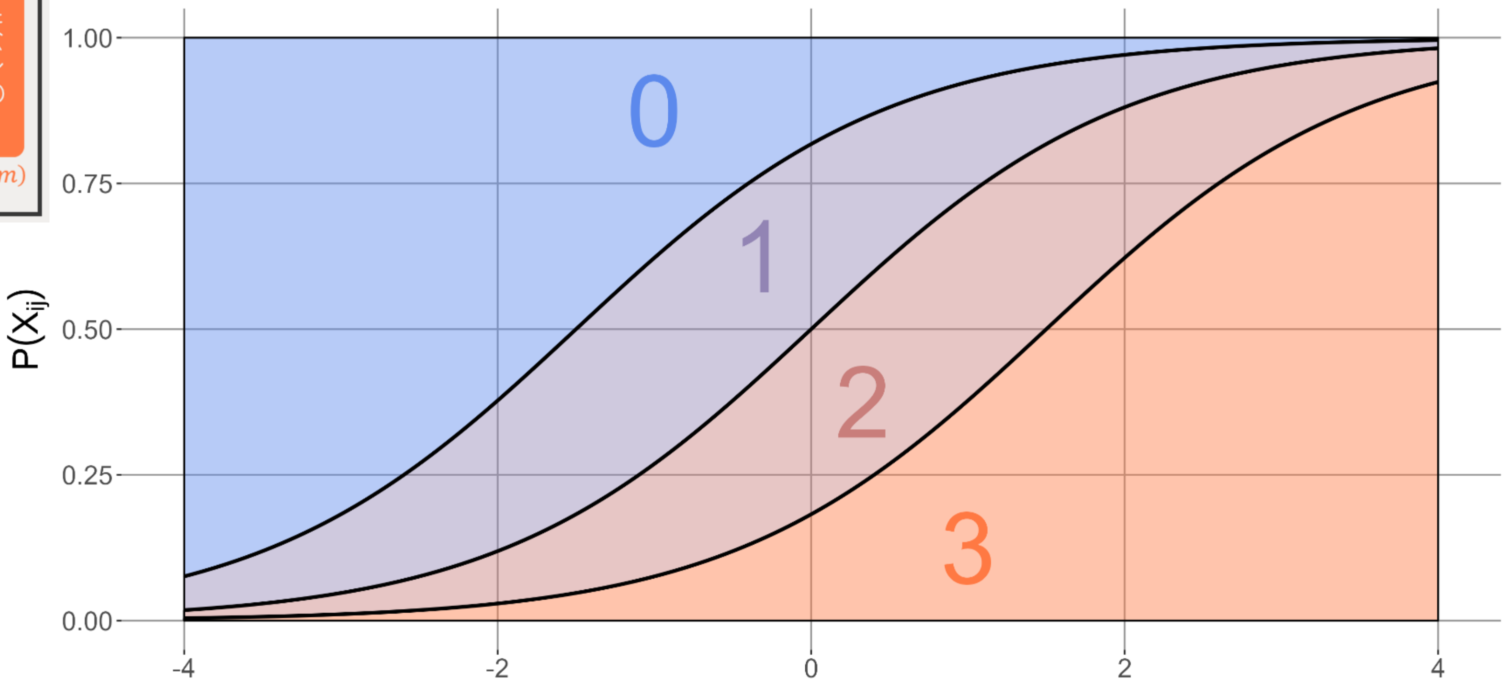
$$P\left(X_{ij}^{(l\_m)} \geq c \middle| \boldsymbol{\eta}_i\right) = \Phi\left[\alpha_j + \beta_j^{(m)} \eta_{id_j^{(m)}} - \beta_j^{(l)} \eta_{id_j^{(l)}} - \tau_c\right]$$

$$P\left(X_{ij}^{(l\_m)} = c \middle| \boldsymbol{\eta}_i\right) = P\left(X_{ij}^{(l\_m)} \geq c \middle| \boldsymbol{\eta}_i\right) - P\left(X_{ij}^{(l\_m)} \geq c + 1 \middle| \boldsymbol{\eta}_i\right)$$



$\mu_j^{(l)} + \beta_j^{(l)} \eta_{id_j^{(l)}}$
が大きい

文 $j^{(l)}$ が強く
選ばれやすい



$\alpha_j + \beta_j^{(m)} \eta_{id_j^{(m)}} - \beta_j^{(l)} \eta_{id_j^{(l)}}$

$\mu_j^{(m)} + \beta_j^{(m)} \eta_{id_j^{(m)}}$
が大きい

文 $j^{(m)}$ が強く
選ばれやすい

一般的なIRTモデルと比較

正規累積モデル

二値

2パラメータ
正規累積モデル

$$P\left(X_{ij}=1\middle|\eta_i\right)=\Phi\left[a_j\left(\eta_i-b_j\right)\right]$$

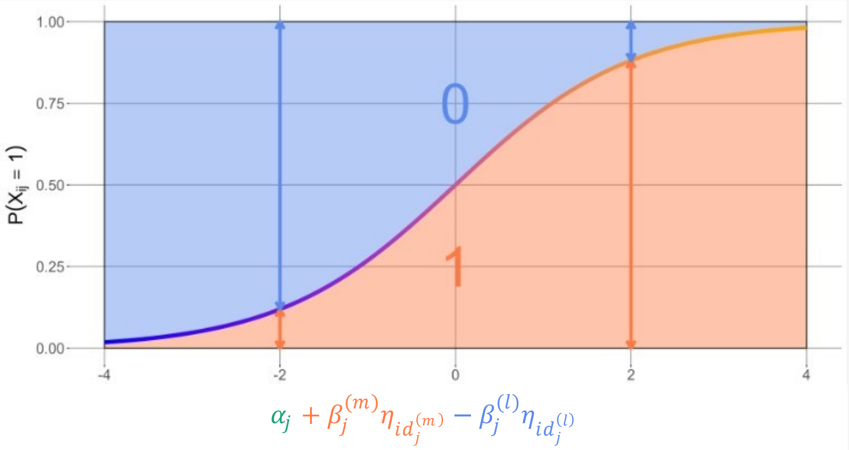
多値

段階反応モデル

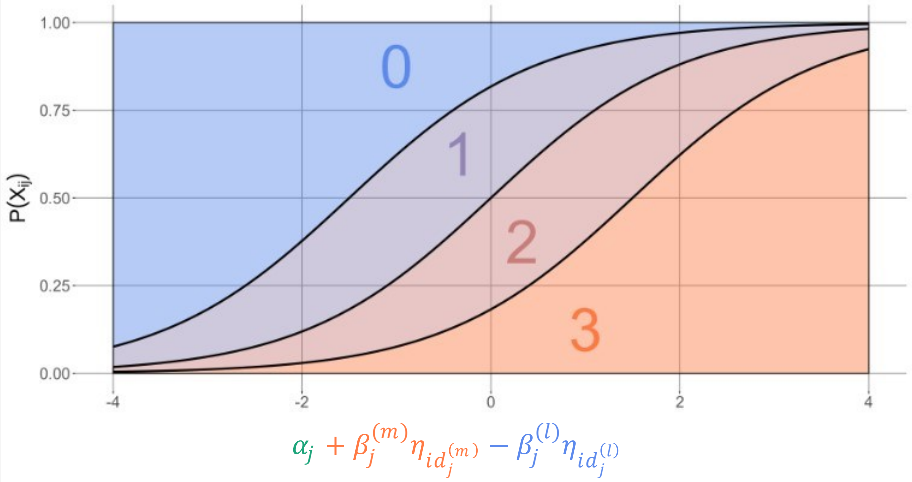
$$P\left(X_{ij}\geq c\middle|\eta_i\right)=\Phi\left[a_j\left(\eta_i-b_j\right)-\tau_c\right]$$
$$P\left(X_{ij}=c\middle|\eta_i\right)=P\left(X_{ij}\geq c\middle|\eta_i\right)-P\left(X_{ij}\geq c+1\middle|\eta_i\right)$$

Thurstonian IRT

$$P\left(X_{ij}^{(l\_m)}=1\middle|\boldsymbol{\eta}_i\right)=\Phi\left[\alpha_j+\beta_j^{(m)}\eta_{id_j^{(m)}}-\beta_j^{(l)}\eta_{id_j^{(l)}}\right]$$



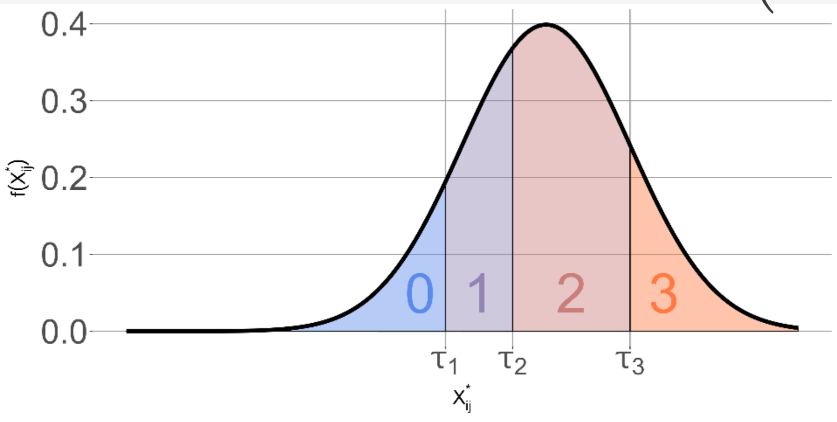
$$P\left(X_{ij}^{(l\_m)}\geq c\middle|\boldsymbol{\eta}_i\right)=\Phi\left[\alpha_j+\beta_j^{(m)}\eta_{id_j^{(m)}}-\beta_j^{(l)}\eta_{id_j^{(l)}}-\tau_c\right]$$
$$P\left(X_{ij}^{(l\_m)}=c\middle|\boldsymbol{\eta}_i\right)=P\left(X_{ij}^{(l\_m)}\geq c\middle|\boldsymbol{\eta}_i\right)-P\left(X_{ij}^{(l\_m)}\geq c+1\middle|\boldsymbol{\eta}_i\right)$$

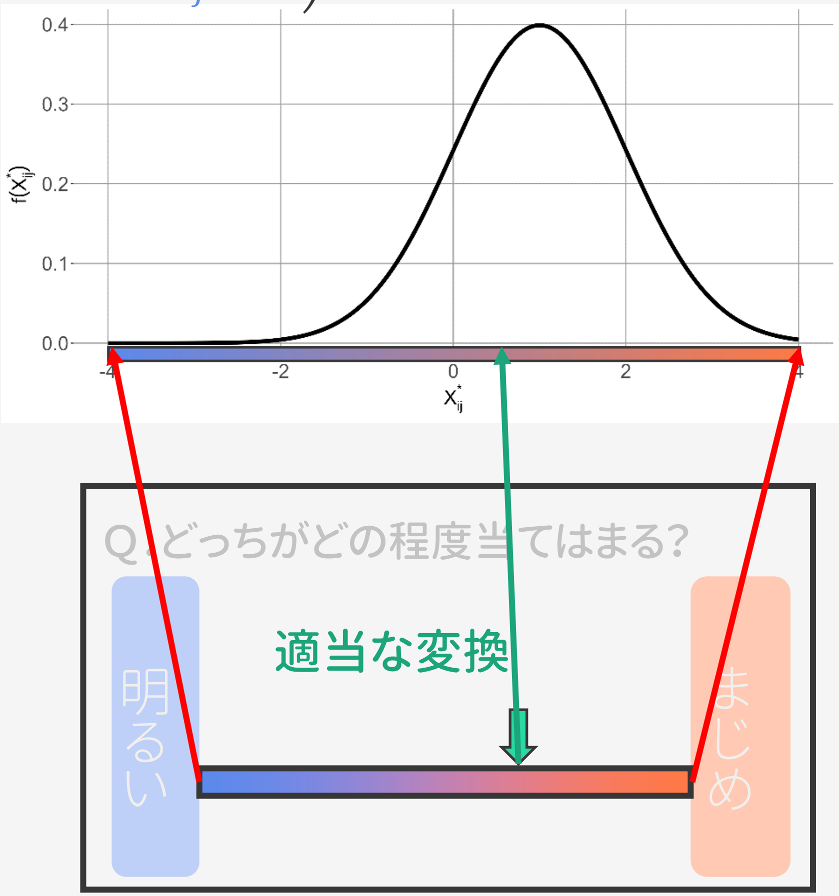


連続反応に対するモデル(Samejima, 1973の拡張)

本来観測したいのはカテゴリではなくその背後の連続量

$$X_{ij}^{*} \sim N\left(\alpha_j + \beta_j^{(m)}\eta_{id_j^{(m)}} - \beta_j^{(l)}\eta_{id_j^{(l)}}, \sigma_\varepsilon^2\right)$$





このまま
考えてみる

$$P\left(X_{ij}^{*} = z \middle| \eta_i\right) = \frac{1}{\sqrt{2\pi\gamma_j^2}} \exp\left(-\frac{\left(\alpha_j + \beta_j^{(m)}\eta_{id_j^{(m)}} - \beta_j^{(l)}\eta_{id_j^{(l)}} - z\right)^2}{2\gamma_j^2}\right)$$

平均 $\alpha_j + \beta_j^{(m)}\eta_{id_j^{(m)}} - \beta_j^{(l)}\eta_{id_j^{(l)}}$,分散γ_j^2の正規分布の確率密度

連続反応モデルのイメージ

平均 $\alpha_j + \beta_j^{(m)}\eta_{id_j^{(m)}} - \beta_j^{(l)}\eta_{id_j^{(l)}}$,分散γ^2の
正規分布の確率密度

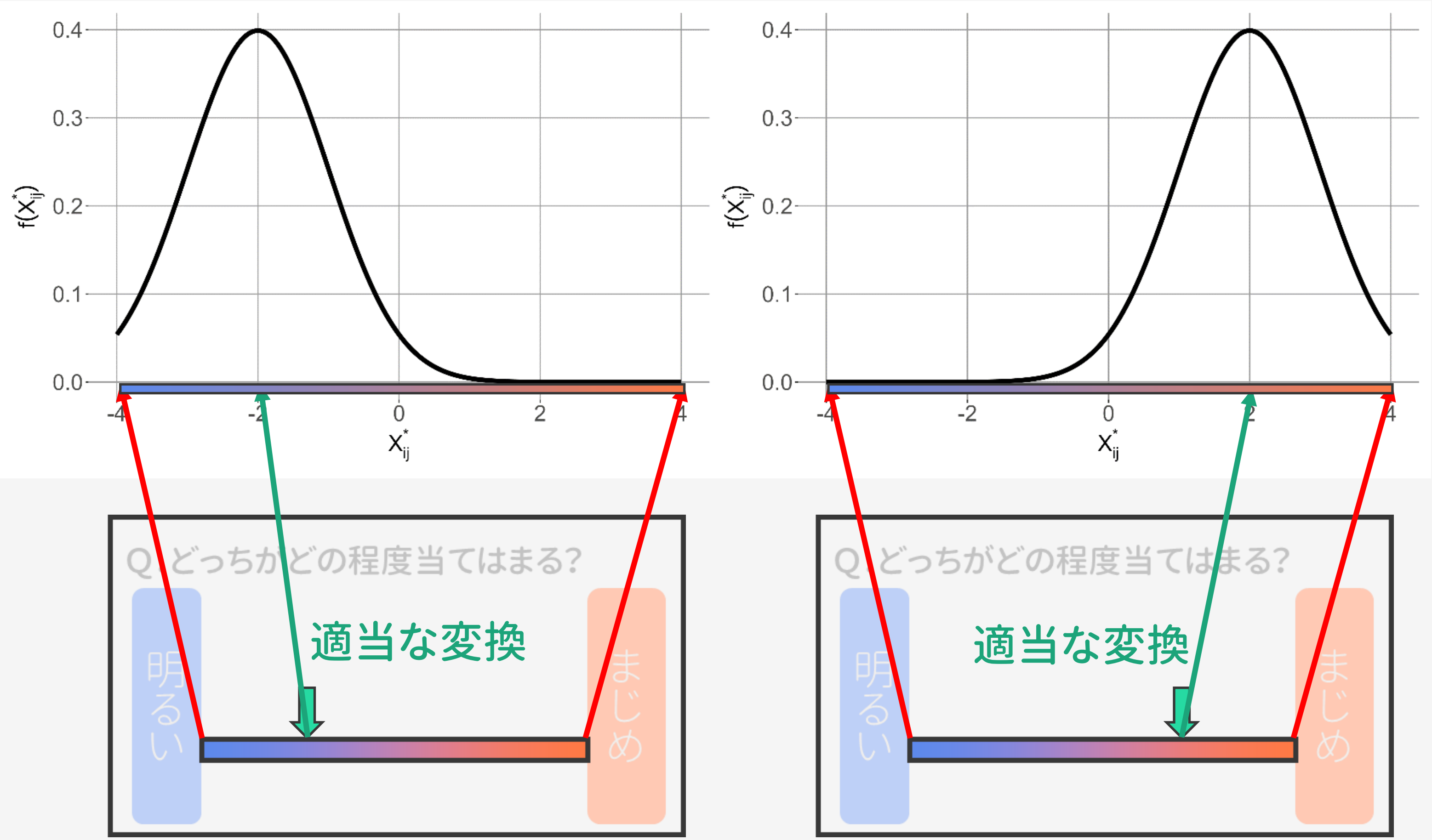
選好の差によって潜在変数の分布が動く

$$X_{ij}^* \sim N\left(\alpha_j + \beta_j^{(m)}\eta_{id_j^{(m)}} - \beta_j^{(l)}\eta_{id_j^{(l)}}, \gamma^2\right)$$

$\mu_j^{(l)} + \beta_j^{(l)}\eta_{id_j^{(l)}}$
が大きい

正規分布の
平均が小さい

文 $j^{(l)}$ が強く
選ばれやすい



$\mu_j^{(m)} + \beta_j^{(m)}\eta_{id_j^{(m)}}$
が大きい

正規分布の
平均が大きい

文 $j^{(m)}$ が強く
選ばれやすい

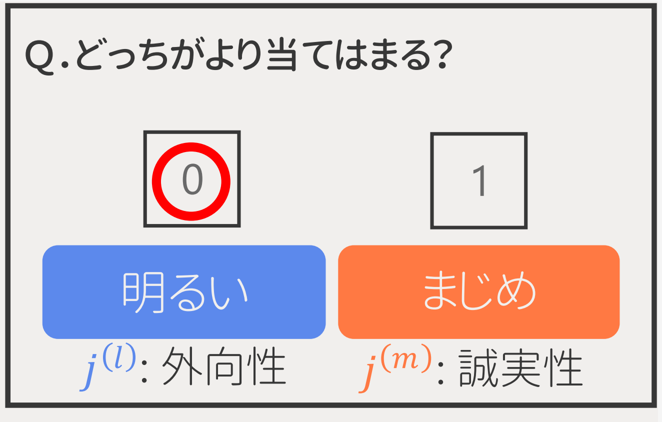
項目情報量

- 測定の目的は回答者の特性値を精度良く推定すること

η_i

- 項目への回答が特性値推定にもたらす情報は?

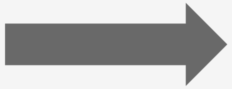
① 二値の場合



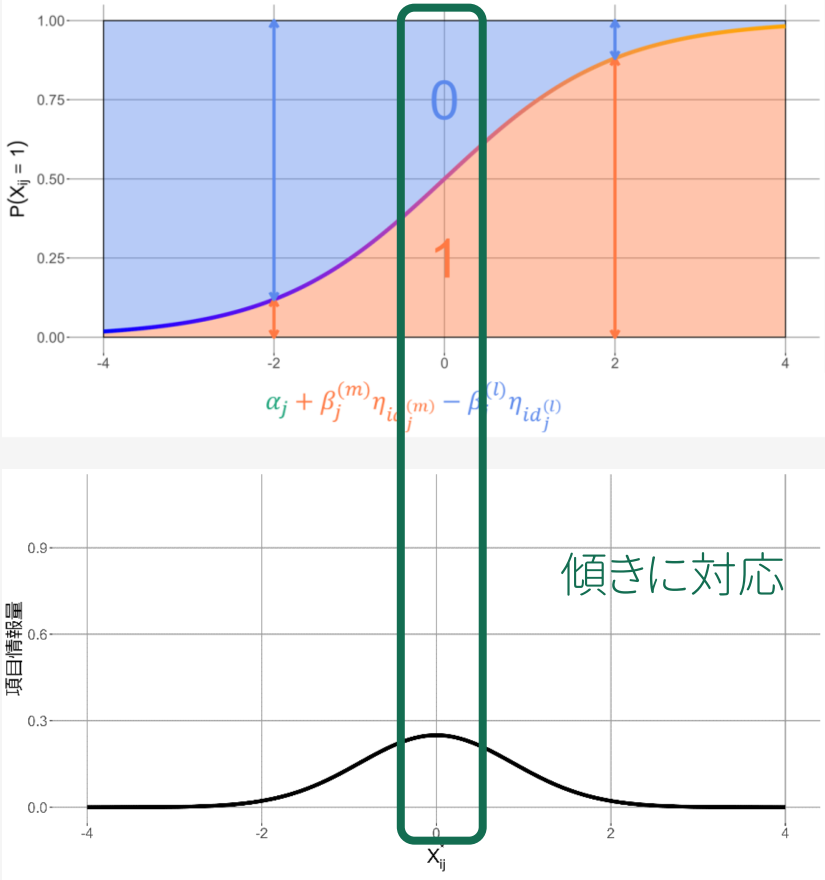
明るい ＞ まじめ

外向性のほうが高い

でもどの程度かはわからない



選好の差が小さくなるような η_i の値なら
情報量が多い項目



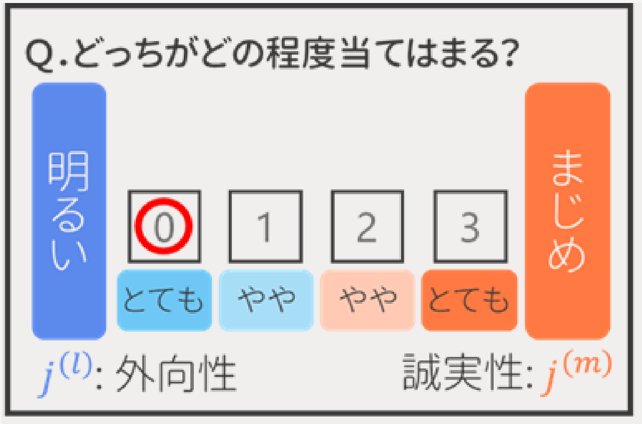
項目情報量

- 測定の目的は回答者の特性値を精度良く推定すること

η_i

- 項目への回答が特性値推定にもたらす情報は?

② 多値の場合

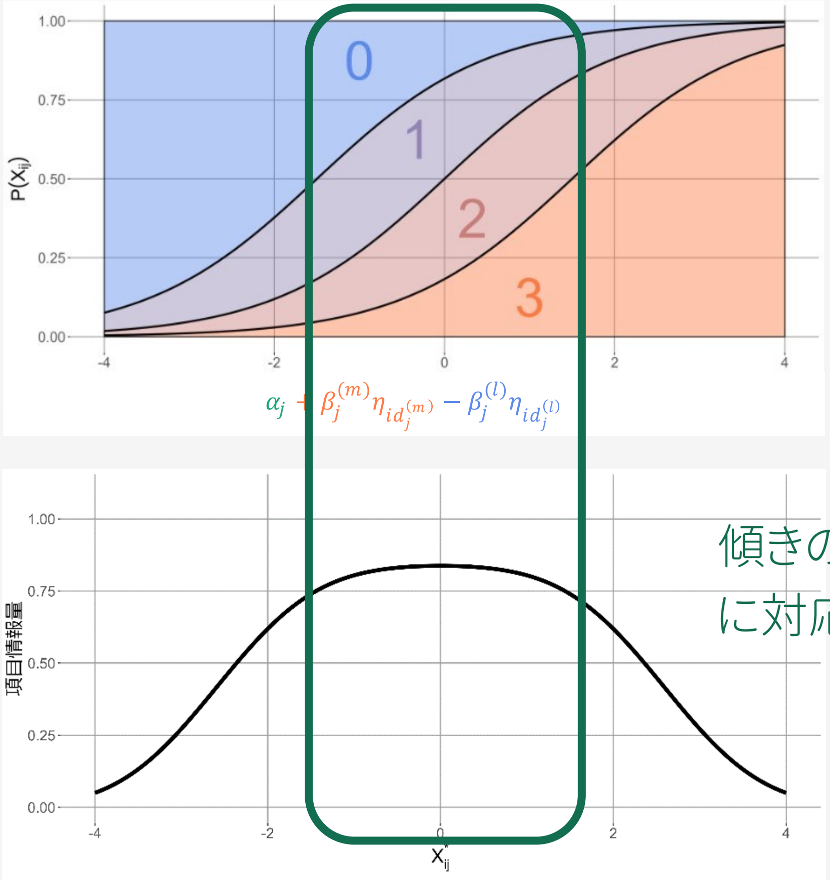


明るい >> まじめ

外向性のほうが高い

二値よりは多少程度がわかる

情報量の多い範囲が
二値よりは広い



傾きの重み付け和
に対応

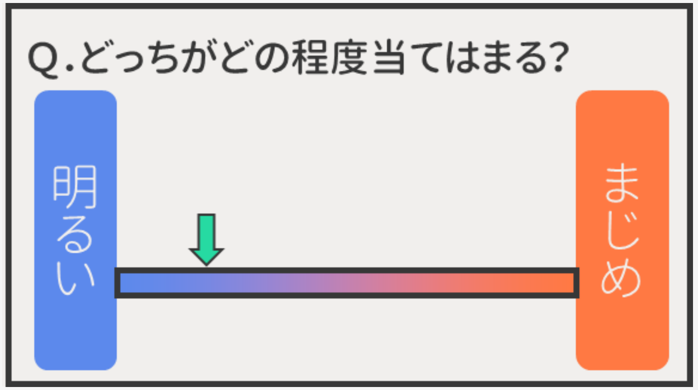
項目情報量

- 測定の目的は回答者の特性値を精度良く推定すること

η_i

- 項目への回答が特性値推定にもたらす情報は?

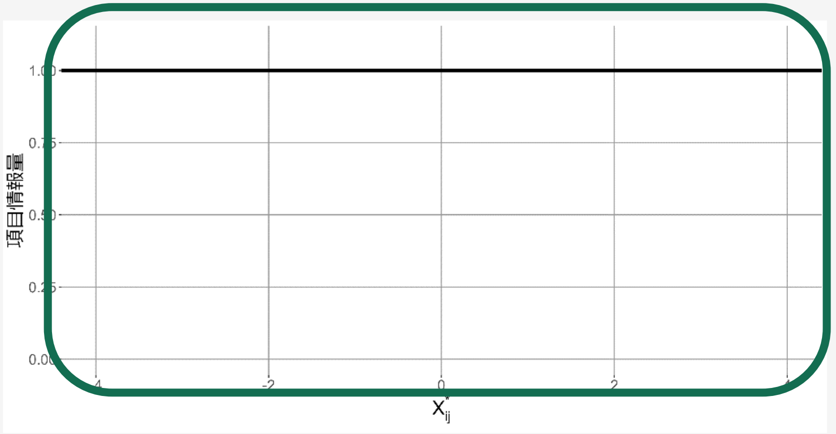
③ 連続の場合



明るい >> まじめ

理論上はどこに回答しても
それが程度を明示している

η_i によらず一定の高い値になる



項目情報量

- 測定の目的は回答者の特性値を精度良く推定すること

$\boldsymbol{\eta}_i$

- 項目情報量＝フィッシャー情報量
  - ▶ 対数尤度の $\boldsymbol{\eta}_i$ に関する二階微分の期待値のマイナス

多次元なら
フィッシャー情報行列

$$\mathbf{I}_j(\boldsymbol{\eta}) = -E\left[\frac{\partial^2}{\partial\boldsymbol{\eta}\partial\boldsymbol{\eta}^\top}\ln P\left(X_{ij}\middle|\boldsymbol{\eta}_i\right)\right]$$

- テスト全体での情報量は単純に和を取れば良い

$$\mathbf{I}(\boldsymbol{\eta}) = \sum \mathbf{I}_j(\boldsymbol{\eta})$$

- すると推定精度が求められる

最尤推定値の共分散行列が$\mathbf{I}(\boldsymbol{\eta})^{-1}$ ➡ 項目情報量は多いほうが嬉しい

カテゴリ化と情報量の関係

具体的な項目情報量の式は

$$I_j(\eta) = \beta_j^2 \sum_{c=1}^{C-1} \frac{\left[P'(X_{ij} \geq c|\eta_i) - P'(X_{ij} \geq c-1|\eta_i)\right]^2}{P(X_{ij} \geq c|\eta_i) - P(X_{ij} \geq c-1|\eta_i)}$$

$= $ 因子負荷量 $\times$ sum(カテゴリごとの情報量)

またの名を**項目識別力**

$P'(X_{ij} \geq c|\eta_i)$は$P(X_{ij} \geq c|\eta_i)$の導関数

▶ 正規累積モデルの場合

$\phi(X_{ij} = c|\eta_i)$

本来特性値は連続量

▶ カテゴリ化することで**情報量が減少してしまう**

情報量の減少の程度は(Bürkner, 2022)

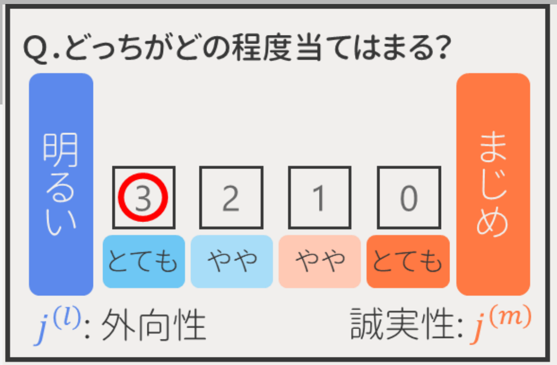
$$0 \leq \sum_{c=1}^{C-1} \frac{\left[P'(X_{ij} \geq c|\eta_i) - P'(X_{ij} \geq c-1|\eta_i)\right]^2}{P(X_{ij} \geq c|\eta_i) - P(X_{ij} \geq c-1|\eta_i)} \leq 1$$

- カテゴリ数が少ないほど小さくなる
- カテゴリ数が無限のとき

($X_{ij}^{(l_m)^*}$が直接測定できるとき)に1になる

▶ 連続反応の場合 $I_j(\eta) = \beta_j^2$

TIRTでの情報量



段階反応モデルにおける項目情報量

各文が測定している因子に対応したベクトル $\boldsymbol{\beta}_j$ を用意する

例|3因子を測定する尺度において文lが因子1, 文mが因子3を測定している場合

$$\boldsymbol{\beta}_j = [\beta_j^{(l)} \quad 0 \quad -\beta_j^{(m)}]$$

後はおなじ $\mathbf{I}_j(\boldsymbol{\eta}) = \boldsymbol{\beta}_j^{\top}\boldsymbol{\beta}_j \sum_{c=1}^{C-1} \frac{\left[P'(X_{ij} \geq c|\eta_i) - P'(X_{ij} \geq c-1|\eta_i)\right]^2}{P(X_{ij} \geq c|\eta_i) - P(X_{ij} \geq c-1|\eta_i)}$

$$= \begin{bmatrix} {\beta_j^{(l)}}^2 & 0 & -\beta_j^{(l)}\beta_j^{(m)} \\ 0 & 0 & 0 \\ -\beta_j^{(l)}\beta_j^{(m)} & 0 & {\beta_j^{(m)}}^2 \end{bmatrix} \sum_{c=1}^{C-1} \frac{\left[P'(X_{ij} \geq c|\eta_i) - P'(X_{ij} \geq c-1|\eta_i)\right]^2}{P(X_{ij} \geq c|\eta_i) - P(X_{ij} \geq c-1|\eta_i)}$$

実験デザイン

参加者

クラウドソーシングで募集した日本人500名(18歳以上)

使用した項目

Big-Five factor marker questionnaire (Apple & Neff, 2012)

5因子×各10文＝計50文

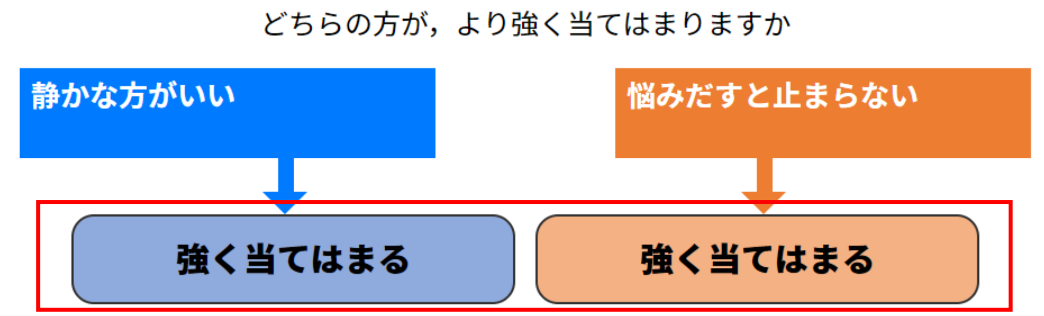
▶ **因子のペアのバランスを取るように**25ペアを構成

	A	B
1	すぐにストレスがたまってしまう	おしゃべりではない
2	他人を気づかうことはない	語彙が豊富である
3	いつもリラックスしていることが多い	想像力が豊かである
4	心配性である	持ち物が整理できないほうだ
5	いつも用意周到である	素晴らしいアイディアを持っている
6	動揺しやすい	盛り上げ役である
7	人を馬鹿にするほうだ	抽象的な考えを理解するのが苦手だ
8	慌てやすい	抽象的な考えには興味がない
9	引っ込み思案である	細かいことに気がつく
10	あまり話すことがない	ものわかりが良いほうだ
11	気分をコロコロ変える	すぐに雑用を済ませる
12	他人に興味がある	整頓するのが好きである
13	気分が著しく変化するほうだ	他の人のために時間を割くほうだ
14	人前でもあがらない	人に共感しやすい
15	無茶なことをする	難しい言葉を使うほうだ
16	落ち込むことはめったにない	自分から話しかけるほうである
17	他人の問題には興味がない	予定に従うほうだ
18	人から注目を浴びるのは好きではない	整理整頓を怠りがち
19	パーティでは色々な人と話すほうだ	優しい心を持っている
20	仕事や学習をさぼることが多い	アイディアが乏しいほうだ
21	イライラしやすい	他人にはまったく興味がない
22	他の人の気持ちがわかる	張り切って仕事や学習に取り組むほうだ
23	落ち込むことが多い	いろんなことを反省しては時間を過ごす
24	人見知りする	人を安心させる
25	注目の的になるのは嫌ではない	アイディアが豊富である

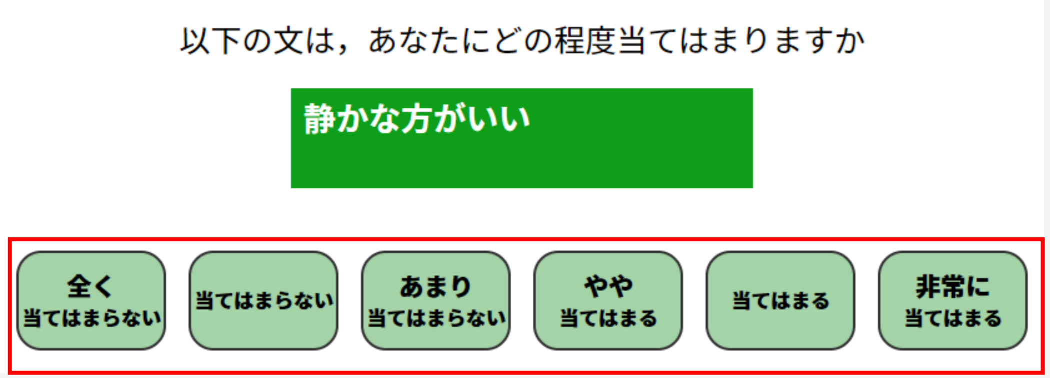
実験デザイン(つづき)

回答フォーマット ▶ 全てクリック1回で回答

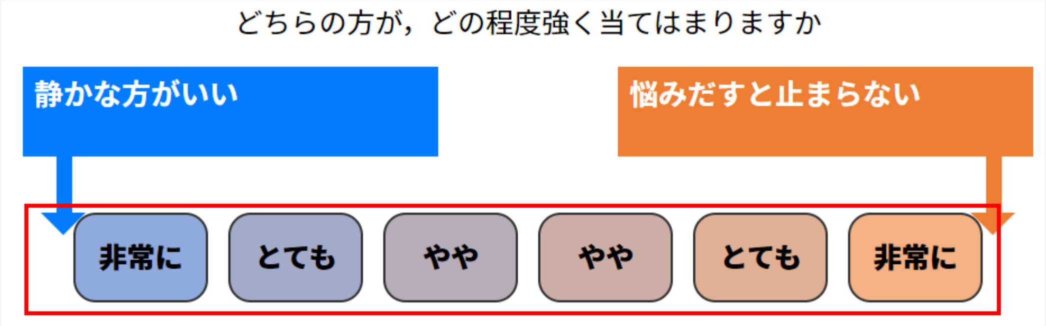
二択



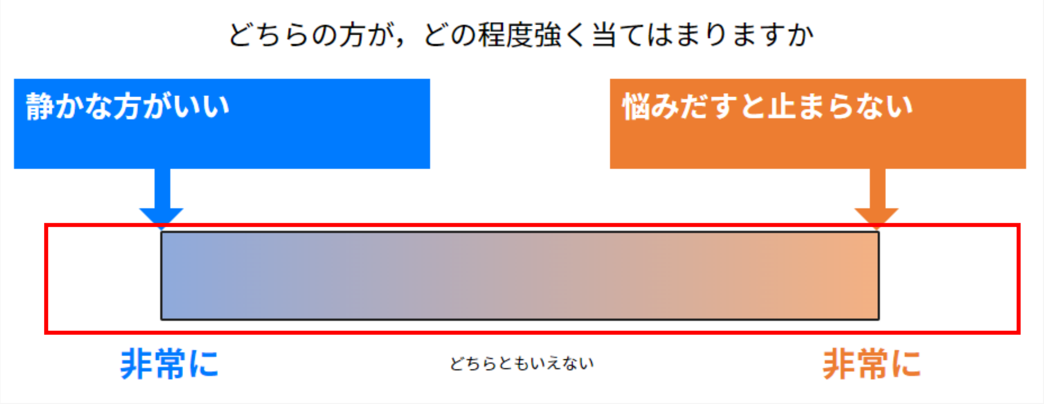
リッカート



多肢選択



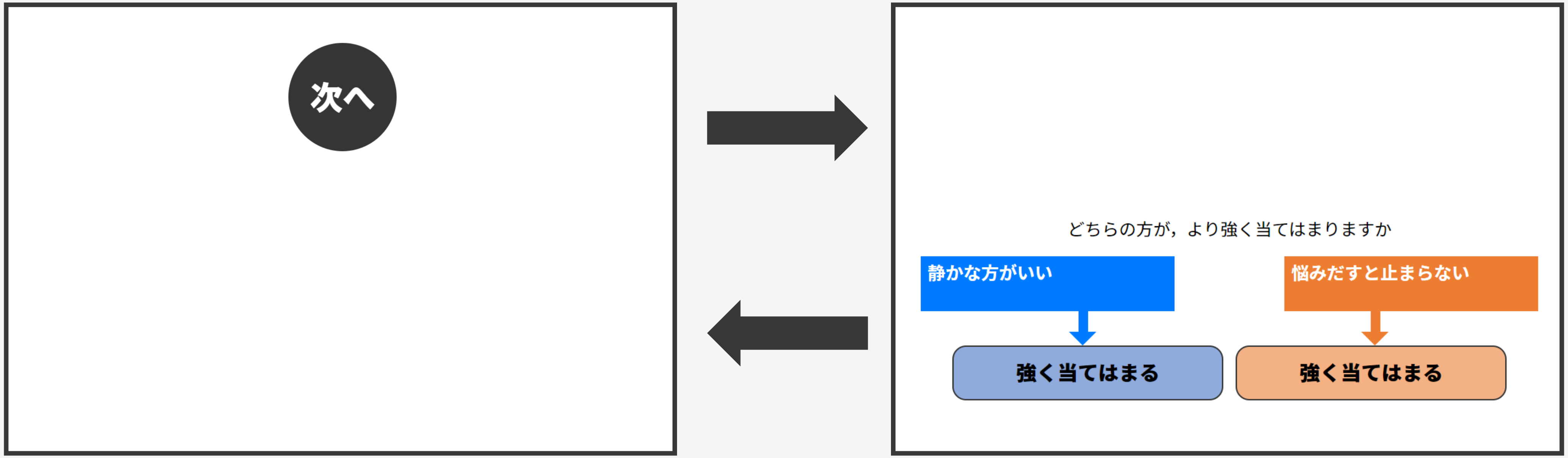
スライダー



実験デザイン(つづき)

各項目の間に注視点のようなものを用意

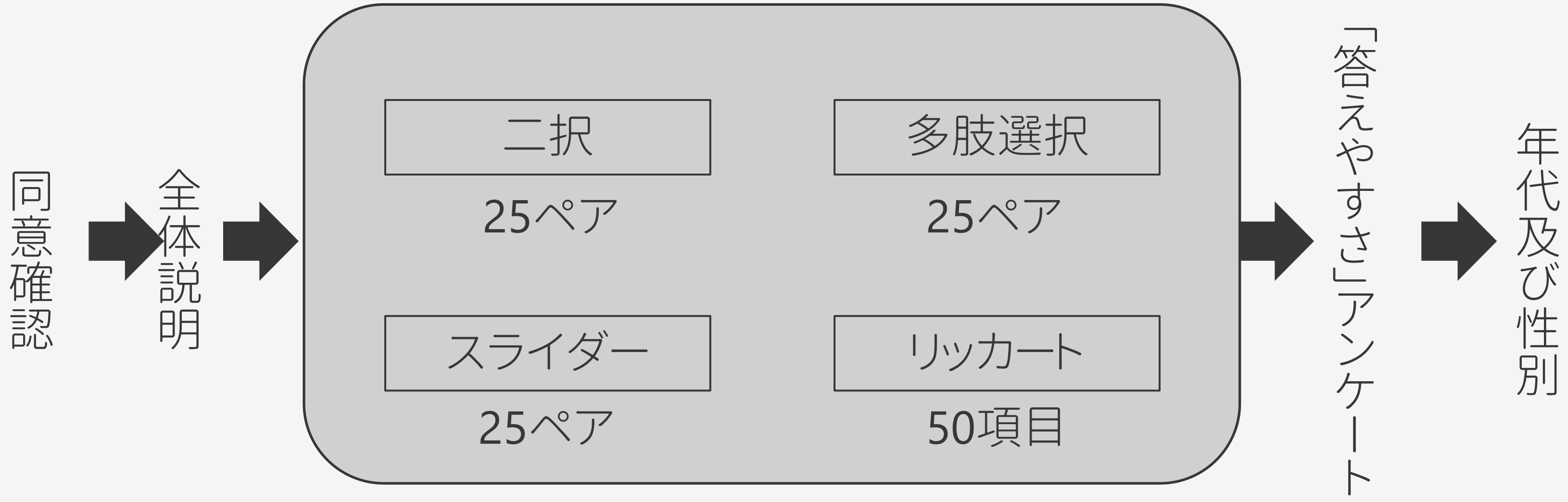
マウスカーソルを乗せると次の項目が出てくる



全てのフォーマットで 次へ から回答ボタンの垂直距離は統一

▶ 回答のために動かすマウスカーソルの距離を統制する

実験の流れ

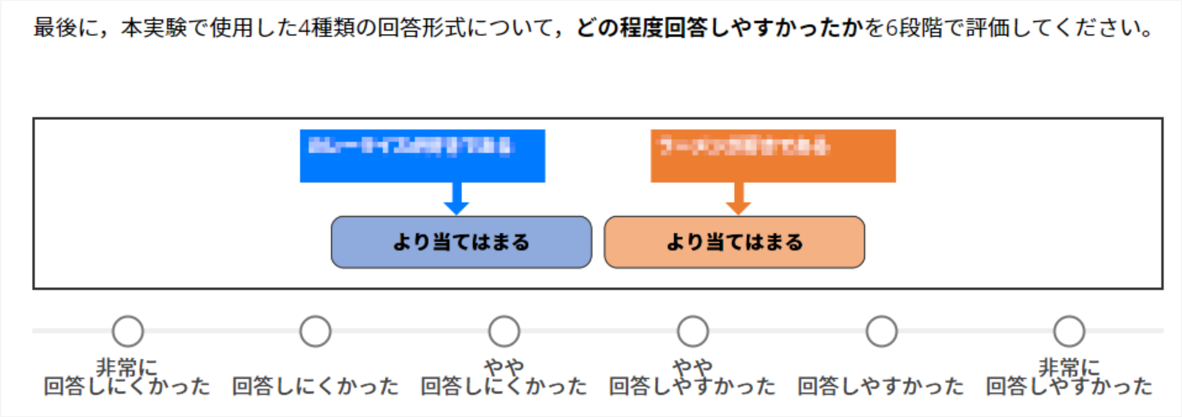


実施順はランダム

フォーマットの実施順，フォーマット内での提示順，ペア内の左右の配置

「答えやすさ」アンケート

6件法



分析モデル

1 二択・多肢選択

$$\alpha_j = \mu_j^{(m)} - \mu_j^{(l)}$$

正規累積
TIRTモデル

$$P\left(X_{ij}^{(l\_m)} \geq c \middle| \boldsymbol{\eta}_i\right) = \Phi\left[\alpha_j + \beta_j^{(m)}\eta_{id_j^{(m)}} - \beta_j^{(l)}\eta_{id_j^{(l)}} - \tau_c\right]$$

$$P\left(X_{ij}^{(l\_m)} = c \middle| \boldsymbol{\eta}_i\right) = P\left(X_{ij}^{(l\_m)} \geq c \middle| \boldsymbol{\eta}_i\right) - P\left(X_{ij}^{(l\_m)} \geq c+1 \middle| \boldsymbol{\eta}_i\right)$$

2 リッカート

正規累積
段階反応モデル

$$P(X_{ij} \geq c | \eta_i) = \left(1 - \Phi\left[\beta_j(\eta_{id_j} - \alpha_j - \tau_c)\right]\right)^{-1}$$

$$P(X_{ij} = c | \eta_i) = P(X_{ij} \geq c | \eta_i) - P(X_{ij} \geq c+1 | \eta_i)$$

3 スライダー

連続反応モデル

Samejima, 1973

$$X_{ij}^* = \log\left(\frac{X_{ij}}{1 - X_{ij}}\right)$$

⬅ 適当な変換

$$P(X_{ij}^* = z | \eta_i) = \frac{1}{\sqrt{2\pi\gamma_j^2}} \exp\left(-\frac{\left(\alpha_j + \beta_j^{(m)}\eta_{id_j^{(m)}} - \beta_j^{(l)}\eta_{id_j^{(l)}} - z\right)^2}{2\gamma_j^2}\right)$$

項目情報量

$I_j(\eta) =$ 因子負荷量 $\times\, \mathrm{sum}($ カテゴリごとの情報量 $)$

テスト情報量は総和＋事前分布の情報量

$$\mathbf{I}(\boldsymbol{\eta}) = \sum \mathbf{I}_j(\boldsymbol{\eta}) + \mathbf{I}_\Sigma$$

1 二択・多肢選択

ロジスティック
TIRTモデル

$$\mathbf{I}_j(\boldsymbol{\eta}) = \boldsymbol{\beta}_j^\top \boldsymbol{\beta}_j \sum_{c=1}^{C-1} \frac{\left[P'(X_{ij} \geq c|\eta_i) - P'(X_{ij} \geq c-1|\eta_i)\right]^2}{P(X_{ij} \geq c|\eta_i) - P(X_{ij} \geq c-1|\eta_i)}$$

2 リッカート

ロジスティック
段階反応モデル

$$I_j(\eta) = \beta_j^2 \sum_{c=1}^{C-1} \frac{\left[P'(X_{ij} \geq c|\eta_i) - P'(X_{ij} \geq c-1|\eta_i)\right]^2}{P(X_{ij} \geq c|\eta_i) - P(X_{ij} \geq c-1|\eta_i)} \Rightarrow \mathbf{I}_j(\boldsymbol{\eta}) = \begin{bmatrix} I_j(\eta) & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix}$$

3 スライダー

連続反応モデル

$$\mathbf{I}_j(\boldsymbol{\eta}) = \boldsymbol{\beta}_j^\top \boldsymbol{\beta}_j$$

Samejima, 1973

分析方針

テスト情報量の比較

分散共分散行列$\mathbf{I}(\boldsymbol{\eta})^{-1}$の対角成分＝推定値の分散を見る　因子単位での情報量→推定精度

情報行列の行列式 $\det(\mathbf{I}(\boldsymbol{\eta}))$ を計算する (D-optimality)　5因子全てまとめての情報量

単位時間あたりテスト情報量の比較

$\det(\mathbf{I}(\boldsymbol{\eta}))$ を合計回答時間T_{sec}（単位：秒）で割った値 $\frac{\det(\mathbf{I}(\boldsymbol{\eta}))}{T_{\mathrm{sec}}}$　5因子全てまとめて

分散共分散行列$\mathbf{I}(\boldsymbol{\eta})^{-1}$の対角成分を$T_{\mathrm{sec}}$倍する $\left[\frac{1}{T_{\mathrm{sec}}}\mathbf{I}(\boldsymbol{\eta})\right]^{-1} = T_{\mathrm{sec}}\mathbf{I}(\boldsymbol{\eta})^{-1}$　因子単位

回答しやすさ

対応のある一要因分散分析

推定の設定

同じ意味のパラメータはモデル間で同じ事前分布を置く

$\alpha_j \sim Normal(0,5)$

$\beta_j \sim student\_t(4,0,2.5)$

$\boldsymbol{\eta}_i \sim MVN(0,\Sigma)$

$\boldsymbol{\Sigma} \sim LKJ(1)$

$\gamma_j \sim Normal(0,5)$

$$P(X^*_{ij} = z|\eta_i) = \frac{1}{\sqrt{2\pi\gamma_j^2}} \exp\left(-\frac{\left(\alpha_j + \beta_j^{(m)}\eta_{id_j^{(m)}} - \beta_j^{(l)}\eta_{id_j^{(l)}} - z\right)^2}{2\gamma_j^2}\right)$$

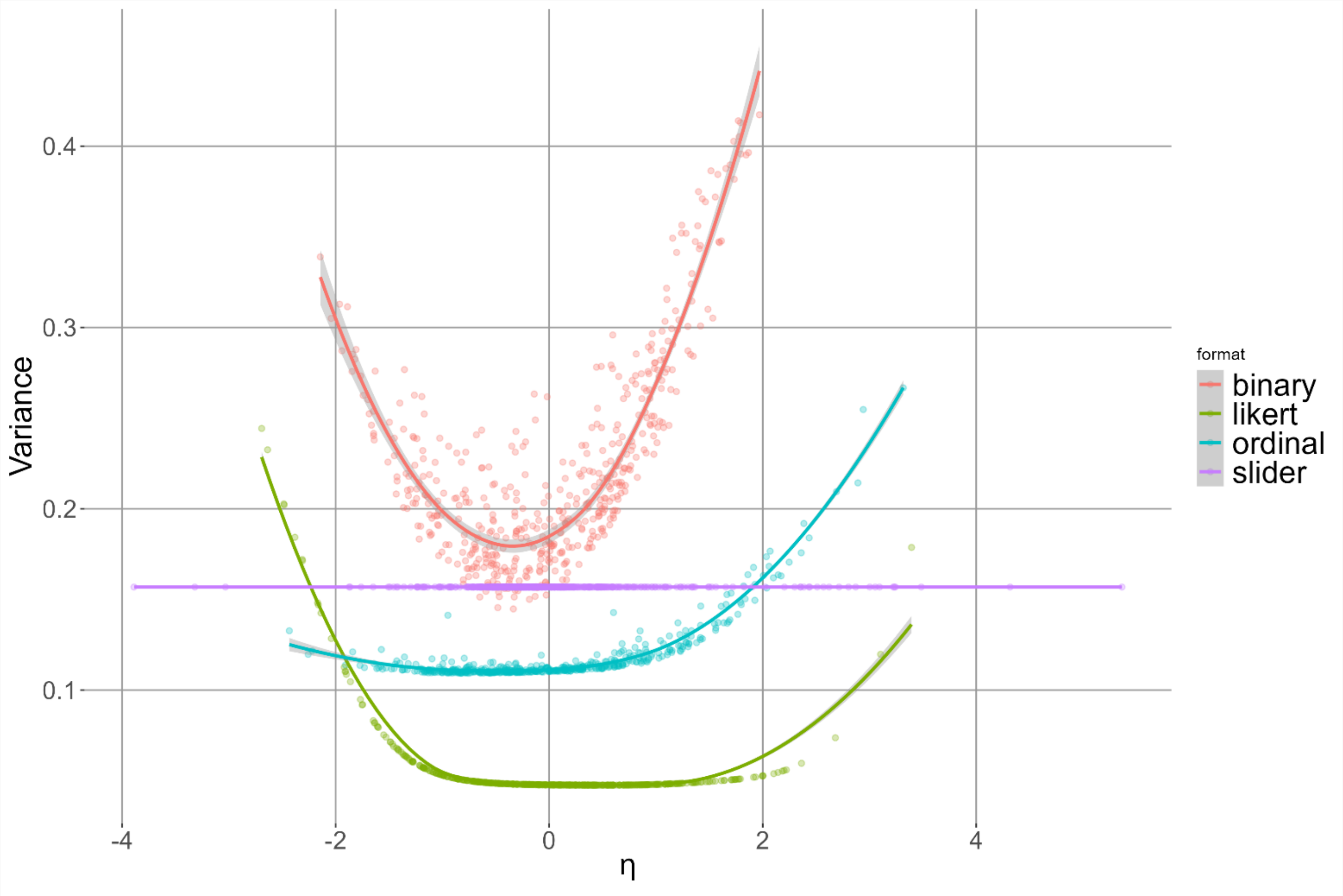
MCMCの設定

`chains=4, n_warmup=150, n_sampling=500`（収束は問題なし）

Outline

テスト情報量

分散共分散行列$\mathbf{I}_j(\boldsymbol{\eta})^{-1}$の(1,1)成分＝因子1の分散のプロット



【前提】

どのフォーマットも各因子は
10項目に出現する
▶ 項目辺り情報量を比べても同じ結果

- リッカートがとても強い
- 二択よりは多肢選択のほうが情報量が多い
- スライダーは中途半端?

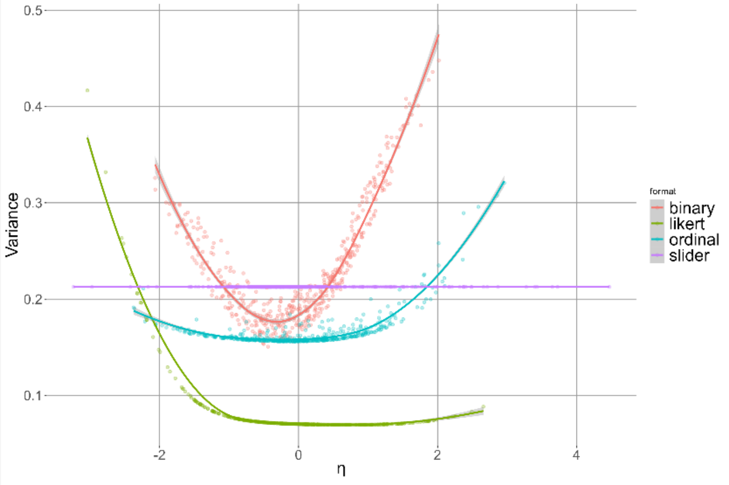
他の因子でも程度の違いはあれど
概ね同じ結果が得られた

テスト情報量

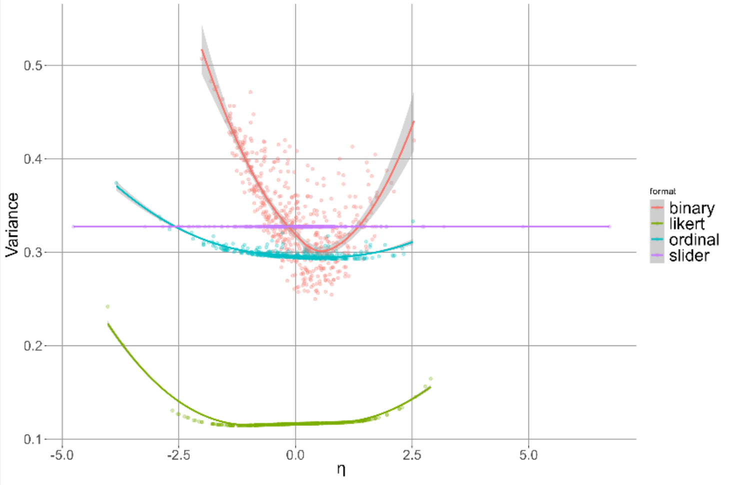
因子2-5の分散のプロット

リッカート＞スライダー≒多肢選択＞二択

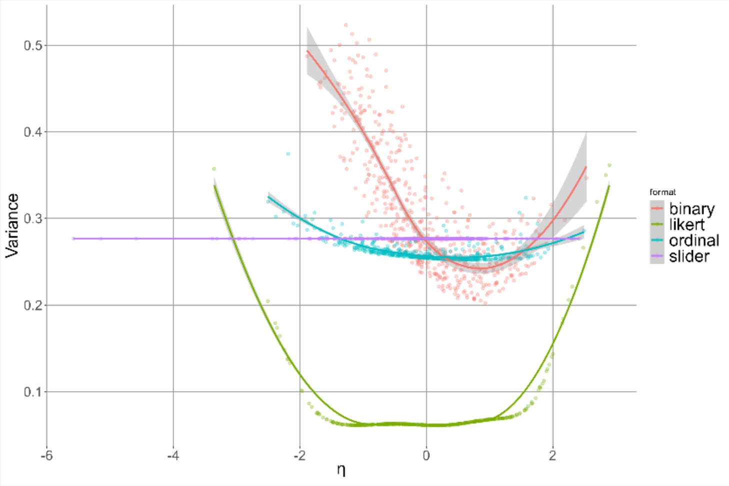
2



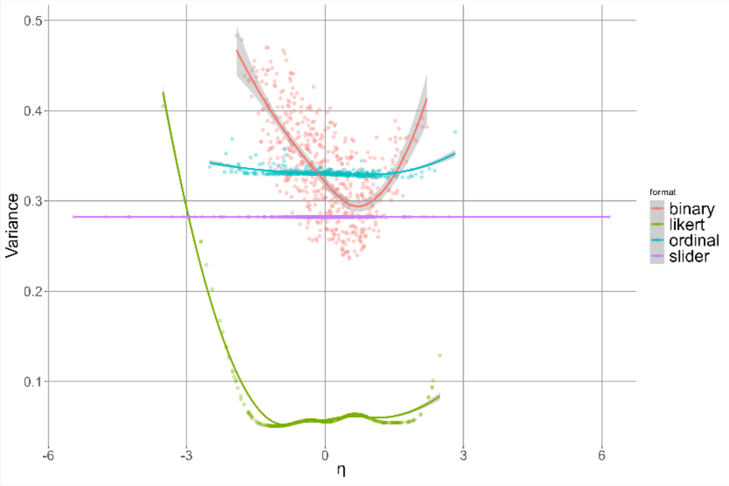
3



4



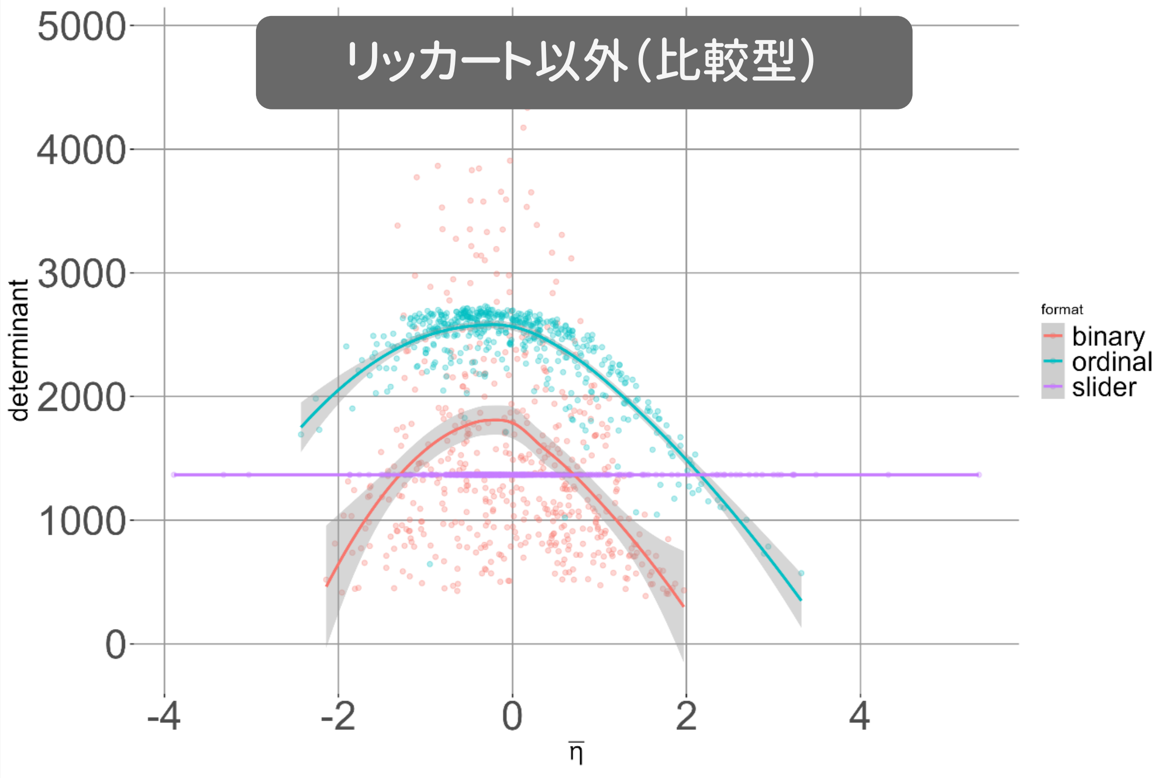
5

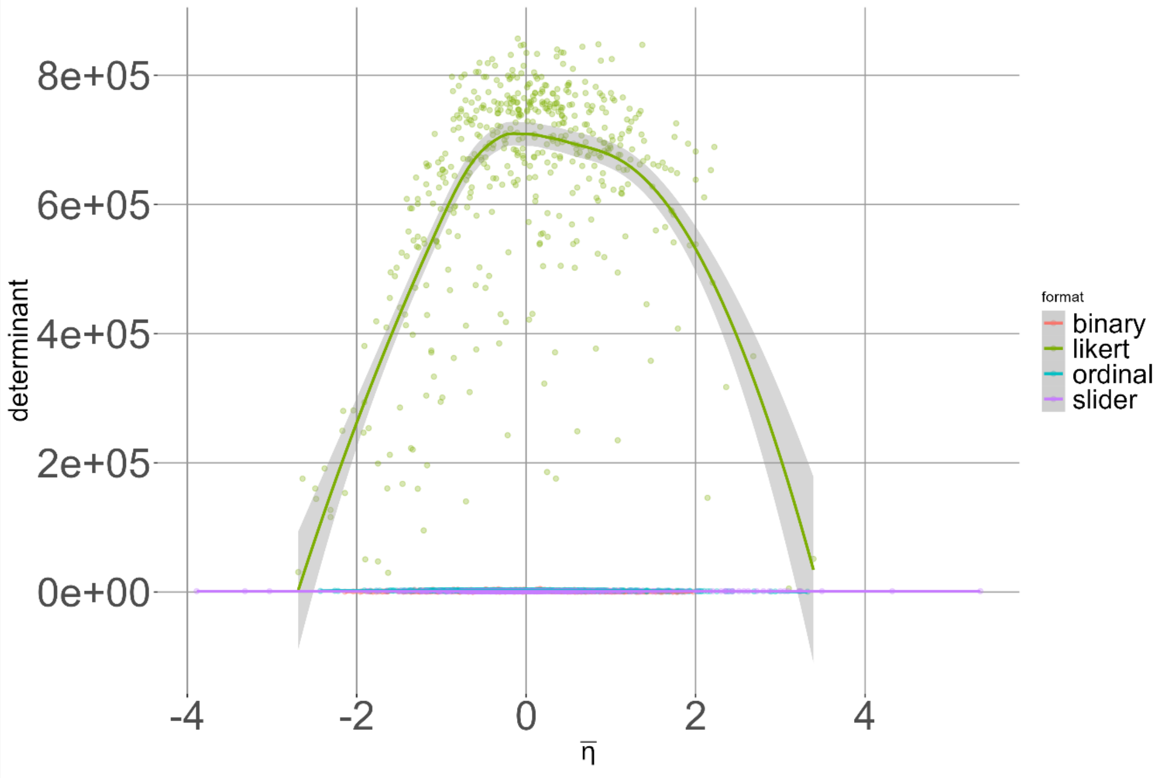


テスト情報行列の行列式

- 個人ごとのばらつきは大きいが
  多肢選択>二択
  スライダーは場所による

- リッカートが大きすぎる

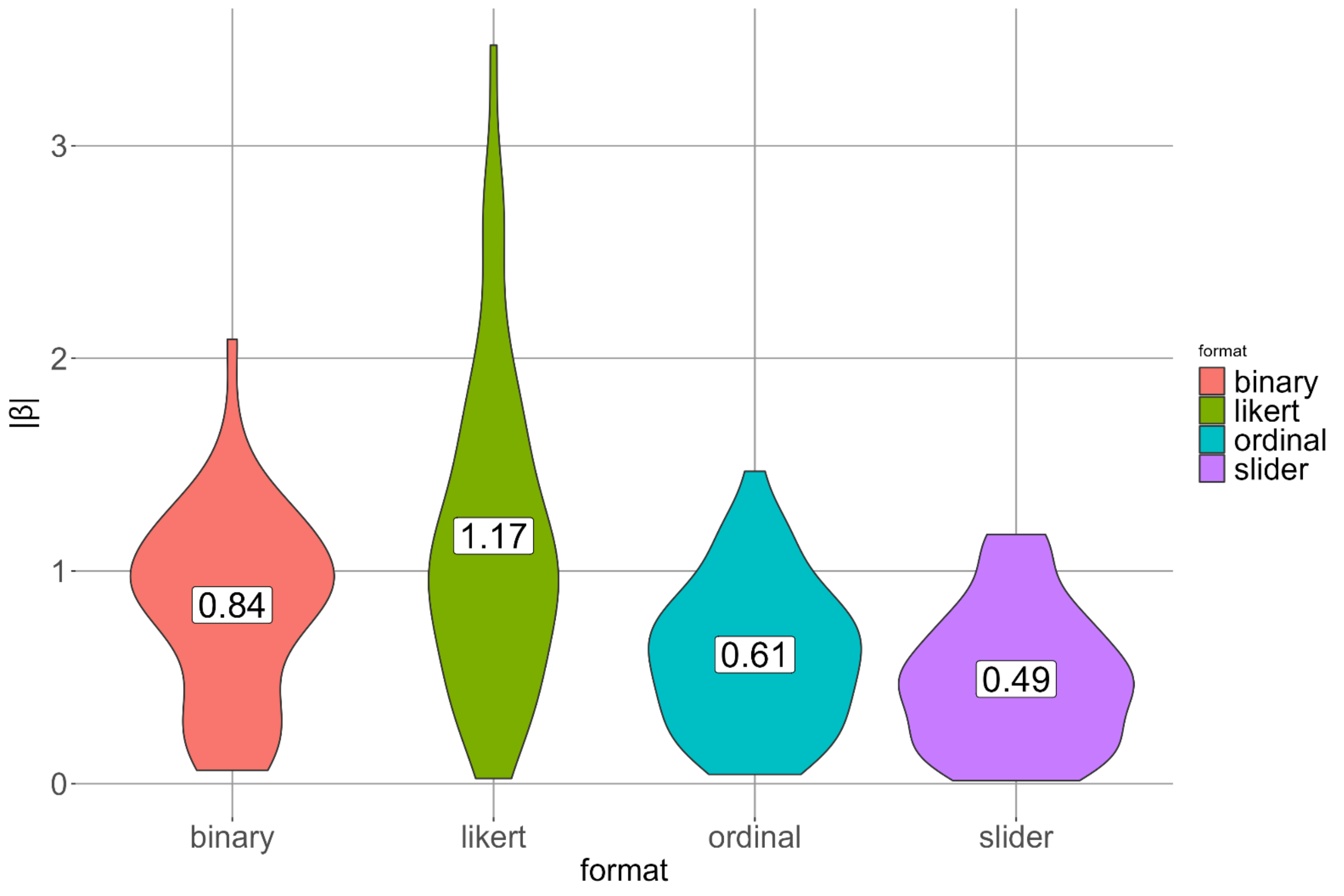




因子負荷と情報量の関係

比較では因子負荷が低くなる

▲ 回答が「もう一方の文」によっても変わるため

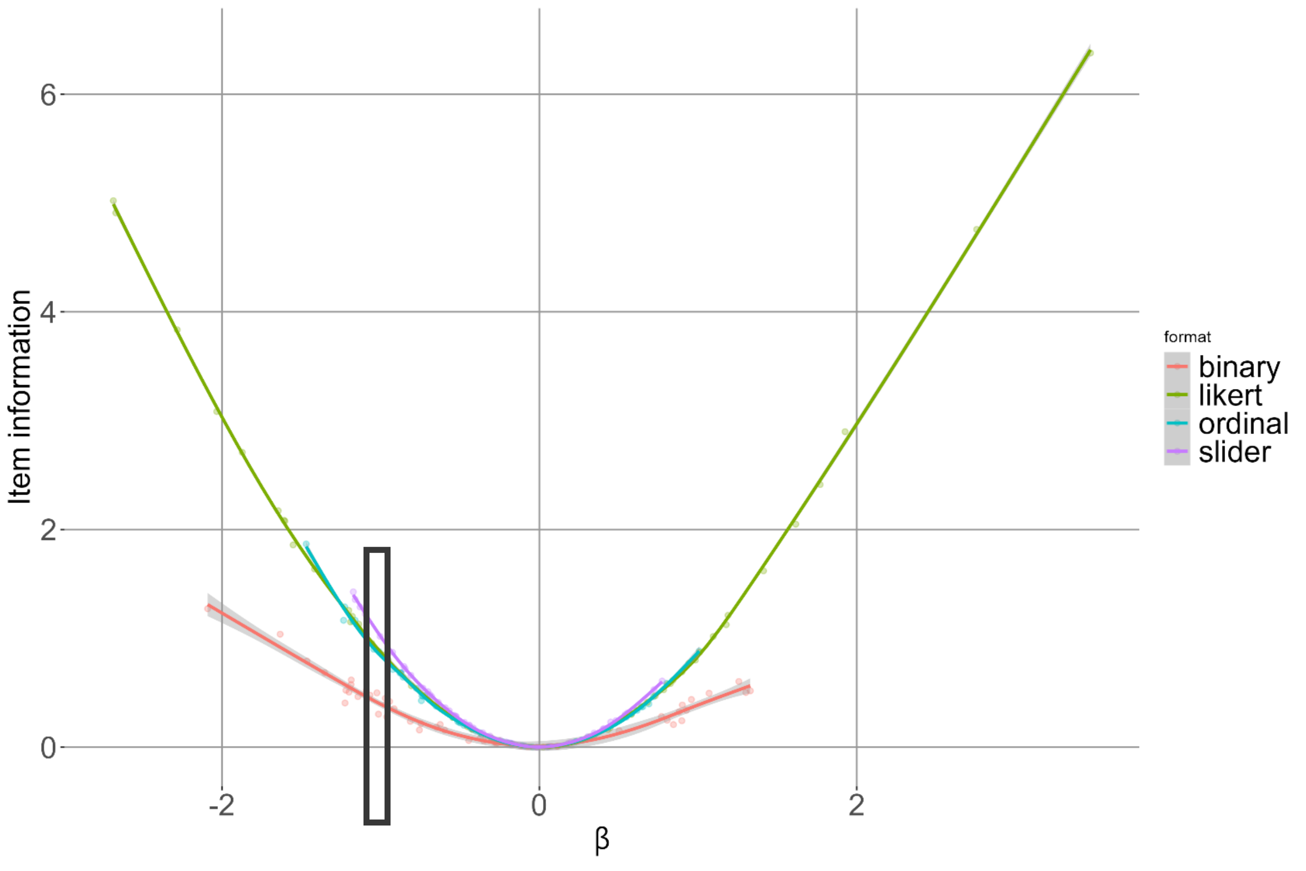


- リッカートではたまに因子負荷の大きな項目が現れる
- 二択のほうが因子負荷は大きい?
  ▶ カテゴリ化によるロスの影響
- スライダーは因子負荷が低くなりがち
  ▶ 原因は要検討

項目識別力と情報量の関係

因子負荷と項目情報量の関係を見てみると

カーブの強さが 二択<リッカート≒多肢選択<スライダー



- スライダーでは因子負荷が小さくても情報量が大きくなりやすい
  ▶ カテゴリ化による損失が少ないため
- リッカートと多肢選択はカテゴリ数が同じなので情報損失の程度も同じ?
- 二択では因子負荷が十分大きくならないと情報量が大きくならない

回答にかかった時間

一問あたり回答時間の差

フォーマット	Mean	Trim10%	SD
二択	3.33	3.04	2.41
多肢選択	3.93	3.61	2.99
スライダー	4.18	3.86	2.76
リッカート	2.57	2.32	2.56

- 比較の場合は選択肢数が少ないほうが速い
  ▶ 多肢選択は時間がかかるが情報量が多い
- リッカートのほうが一問あたりは速い

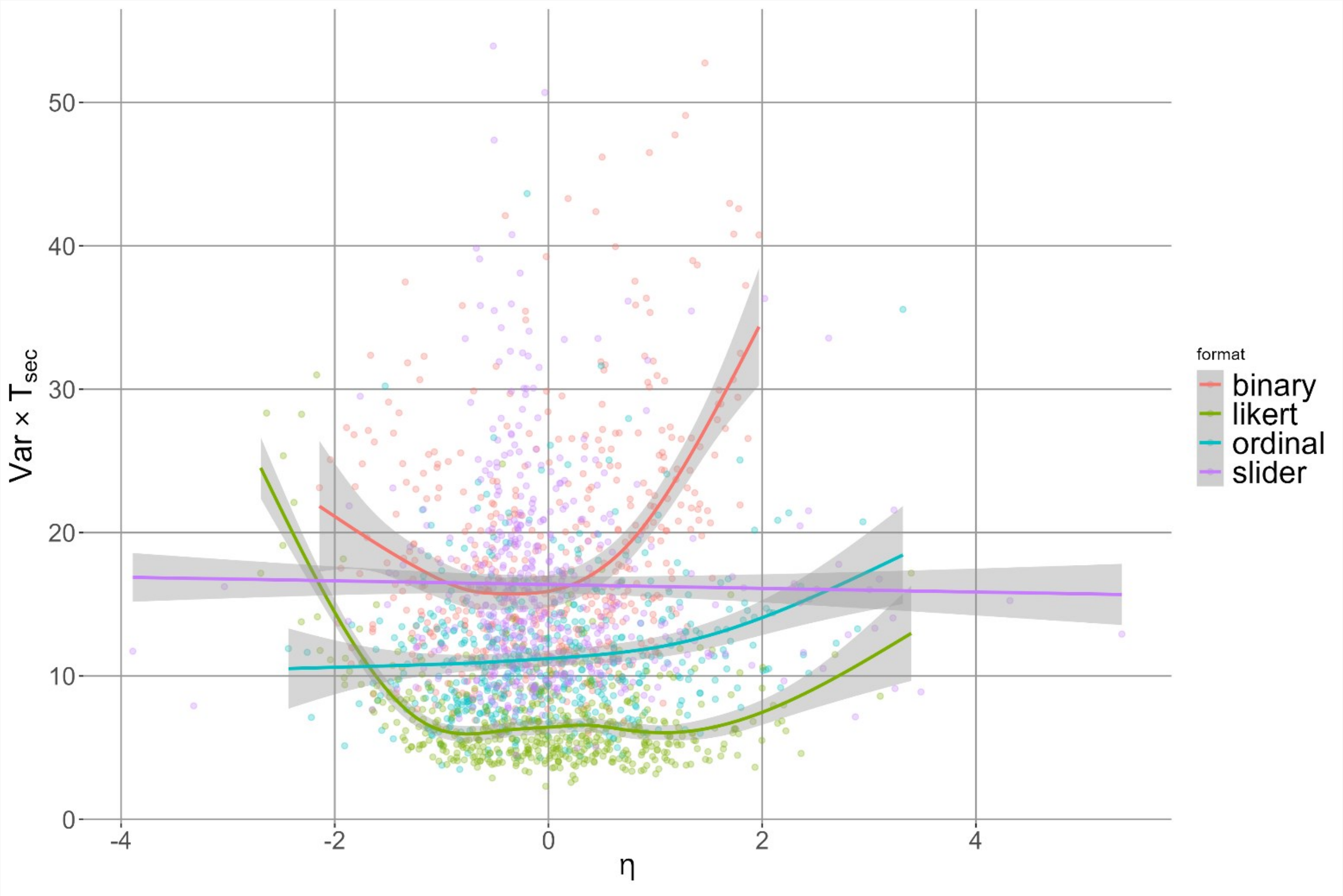
テスト全体での回答時間の差

	フォーマット	Mean	Trim10%	SD
25問	二択	83.26	80.62	32.72
	多肢選択	98.32	94.94	40.41
	スライダー	104.45	100.41	44.71
50問	リッカート	128.67	123.86	49.81

- 比較の場合は選択肢数が少ないほうが速い
- リッカートは項目数が倍なので遅い

テスト情報量

分散共分散行列$\mathbf{I}_j(\eta)^{-1}$の(1,1)成分＝因子1の分散×合計回答時間のプロット



回答時間の違いは
優劣を覆すほどではない

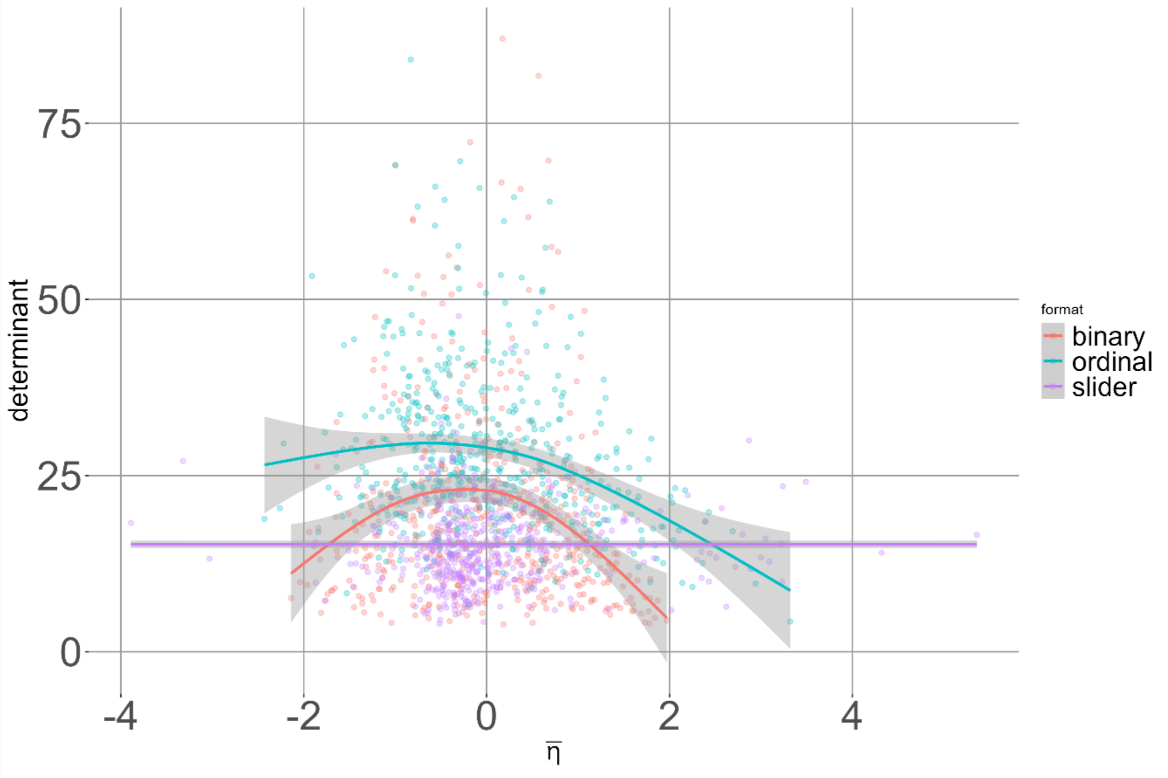
> 他の因子でも程度の違いはあれど
> 概ね同じ結果が得られた

結果

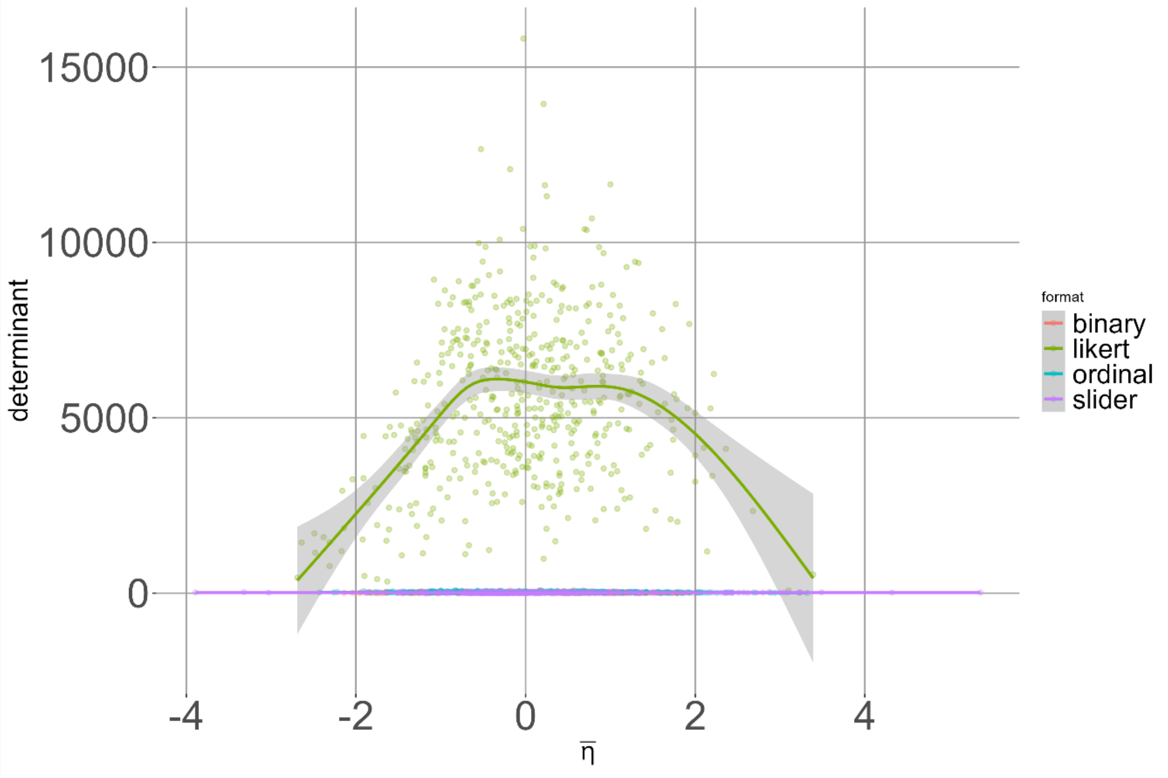
テスト情報行列の行列式

合計回答時間(秒)で割った値

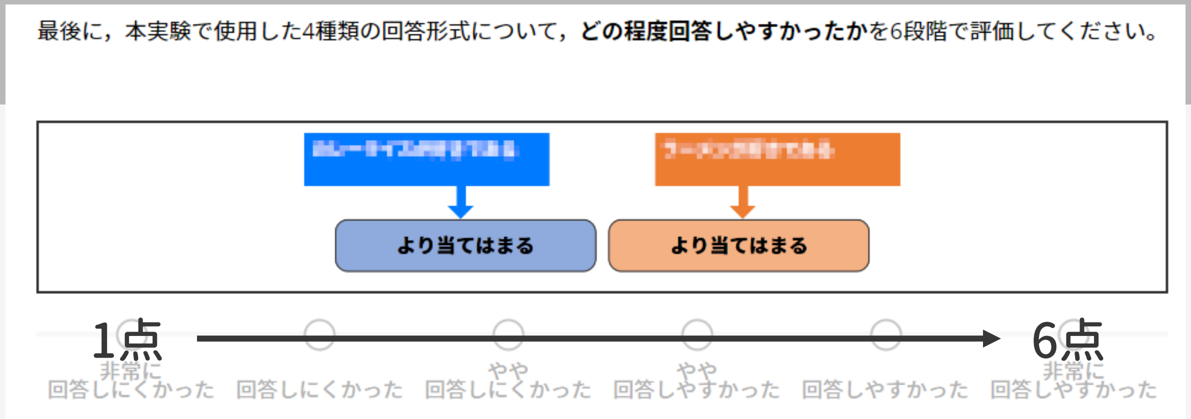
回答時間の個人差が大きいので不安定



結局リッカートが大きすぎる



回答しやすさ



単純集計

フォーマット	Mean	SD
二択	3.48	1.38
多肢選択	3.64	1.17
スライダー	3.62	1.29
リッカート	4.69	0.96

分散分析

要因	SS	df	MS	F	p
個人差	928.2	499	1.86		
フォーマット	471.5	3	157.16	117.56	>.001
残差	2001.3	1497	1.34		
合計	3401.0	1999	1.70		

多重比較

比較型項目は
フォーマットに関係なく
リッカートより答えにくい

比較A	比較B	t	df	p	調整p
二択	多肢選択	2.33	499	.020	.061
	スライダー	1.83	499	.068	.136
	リッカート	15.68	499	>.001	>.001
多肢選択	スライダー	0.32	499	.748	.748
	リッカート	16.23	499	>.001	>.001
スライダー	リッカート	14.28	499	>.001	>.001

まとめ

異なるフォーマットにおける情報量を比較した

二択よりは多肢選択のほうが情報量が多かった(予想通り)

リッカートは強制選択よりも一項目あたり情報量がかなり多い

▲ **強制選択では識別力が低くなってしまうことの影響が大きい**

スライダーはカテゴリ化の損失が少ない一方で識別力が低くなる

▲モデルのせい?回答の敏感性によるノイズのせい?慣れていないせい?

比較型項目の存在意義

多少の情報量を犠牲にしても

- フェイキングなどを抑制できる
- 異なる因子の個人内比較の情報が得られる(非対角成分)
- 意図的に認知負荷を上げることができる?

まとめ（つづき）

回答の困難さのフォーマットによる違いを確認した

比較をさせる時点でリッカートより一段と難しくなる

▶ **比較させるならばどのフォーマットでも困難さはあまり変わらない**

主観的には二択が最も答えにくいが，回答時間は二択が最も速い

▶ **いずれにせよ情報量の違いを覆すほどの差はない**

比較型項目を使うなら

カテゴリ数は人間が識別できる程度（7-9くらい?）に多めにするのが良さそう

※推定するパラメータの増加の影響などは要検討

研究の限界

オンライン実験なのでどの程度真面目に回答してくれたか…