

【社会的望ましさバイアスに頑健な比較型心理測定法の現在】

比較型測定の 項目フォーマットに関する若干の検討

分寺 杏介

神戸大学 経営学研究科



bunji@bear.kobe-u.ac.jp

■ 答えるのが大変

例: OPQ32

(職務行動に関するアセスメントツールとして最も有名なものの一つ)

	Most	Least
I like to discuss abstract concepts	☒	☐
I enjoy interpreting statistics	☒	☐
I feel that people are honest	☒	☐

ざっくり日本語訳

	Most	Least
抽象的な概念について話し合うのが好きだ	☒	☐
統計データを解釈するのは楽しい	☒	☐
人はみな正直であると思う	☒	☐

普通のリッカート尺度だとしたら

抽象的な概念について話し合うのが好きだ

Tourangeau et al.(2000)
のフレームワーク

理解

まずは文章の意味を理解して

想起

過去の記憶などを思い出したりして

判断

どの程度当てはまるかを考えて

回答

選択肢に落とし込む



「ややあてはまる」

くらいかなあ

■ 答えるのが大変

例: OPQ32

(職務行動に関するアセスメントツールとして最も有名なものの一つ)

	Most	Least
I like to discuss abstract concepts	☒	☐
I enjoy interpreting statistics	☒	☐
I feel that people are honest	☒	☐

ざっくり日本語訳

	Most	Least
抽象的な概念について話し合うのが好きだ	☒	☐
統計データを解釈するのは楽しい	☒	☐
人はみな正直であると思う	☒	☐

多肢強制選択型測定の場合

抽象的な概念について話し合うのが好きだ

統計データを解釈するのは楽しい

人はみな正直であると思う

理解

理解

理解

想起

想起

想起

判断

判断

判断

比較判断

回答

統計データを解釈するのは楽しい

が一番かなあ

別の文に関する
記憶の保持なども必要
(Sass et al., 2020)



■ 項目の提示方法などを変えたときの統計的性質を検討する

小さいかもしれないが重要かもしれない話
(本当は比較型測定に限らず重要な話)

■ 以下の2つの研究について報告します

1. 回答フォーマットと項目情報量の関係について

答え方を変えることで、回答時間と得られる情報量はどう変わるのか？

2. 提示する項目要素の長さが再テスト信頼性と答えやすさに与える影響について

形容詞か文か, によって認知負荷が変わるのではないか？

事前登録済み, ただし今回は速報的に報告します。

日本行動計量学会第52回大会で報告した内容の短縮版です。
発表時のスライド(モデルの詳細等)はこちらから▶
ホームページ・ResearchmapからもDL可能です。



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



¹bunji@bear.kobe-u.ac.jp

■ いくつかのフォーマットが考えられる

Q.どっちがより当てはまる?

☒

☐

明るい

まじめ

■ 二択

答えやすい?

Q.どっちがどの程度当てはまる?

明るい

☒

☐

☐

☐

まじめ

とても

やや

やや

とても

■ 多肢選択

情報が多い?

Q.どっちがどの程度当てはまる?

明るい

まじめ

■ スライダースケール

情報をもっと多い?

■ 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...

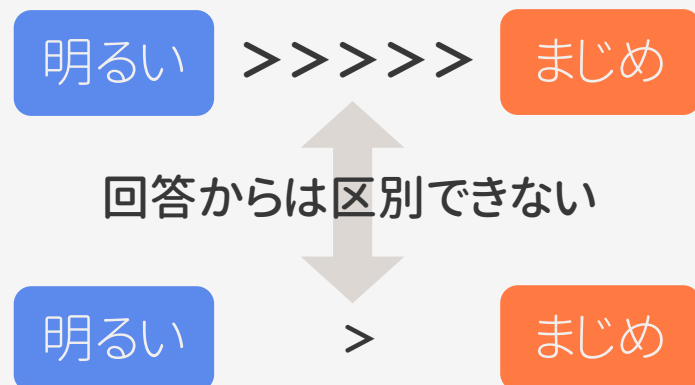
Q.どっちがより当てはまる?

☒

明るい

☐

まじめ



カテゴリ数が多いほど
情報量が多い

Q.どっちがどの程度当てはまる?

明るい

☒

とても

☐

やや

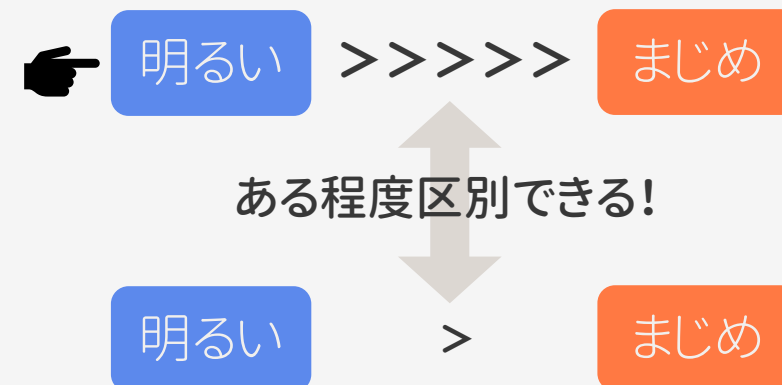
☐

やや

☐

とても

まじめ



■ 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？

■ 強制選択のフォーマット

Q.どっちがより当てはまる?

☒

明るい

☐

まじめ

■ 二択

Q.どっちがどの程度当てはまる?

明るい

☒
☐
☐
☐

まじめ

とても
やや
やや
とても

■ 多肢選択

Q.どっちがどの程度当てはまる?

明るい

↓

まじめ

■ スライダースケール

理論的には

情報量

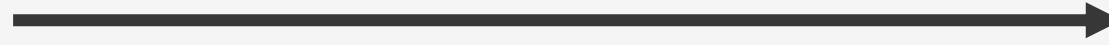
少ない



多い

答えやすさ

答えやすい?



答えにくい?

回答時間

短い?

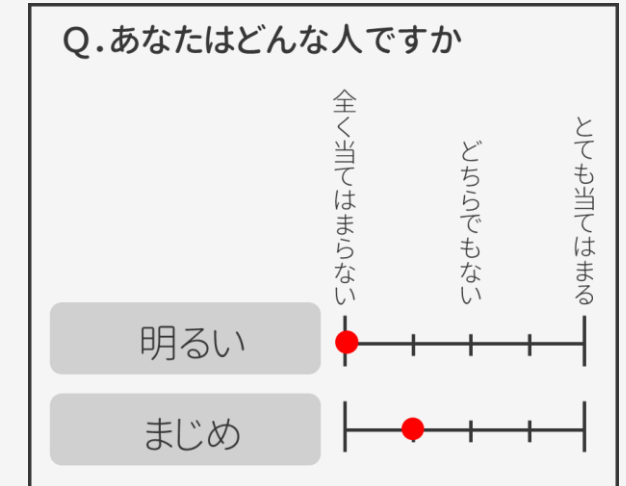


長い?

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

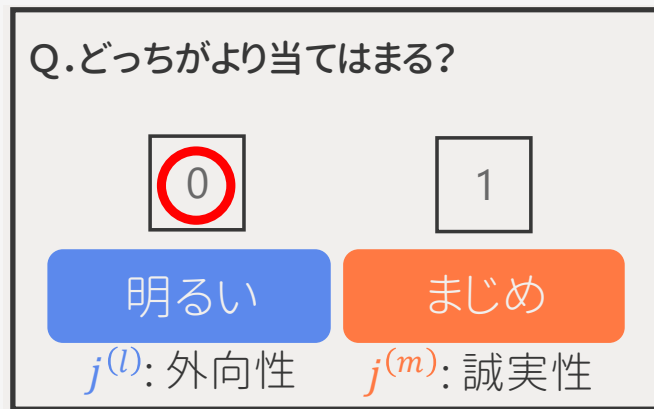


- 測定の目的は回答者の特性値を精度良く推定すること

η_i

- 項目への回答が特性値推定にもたらす情報は？

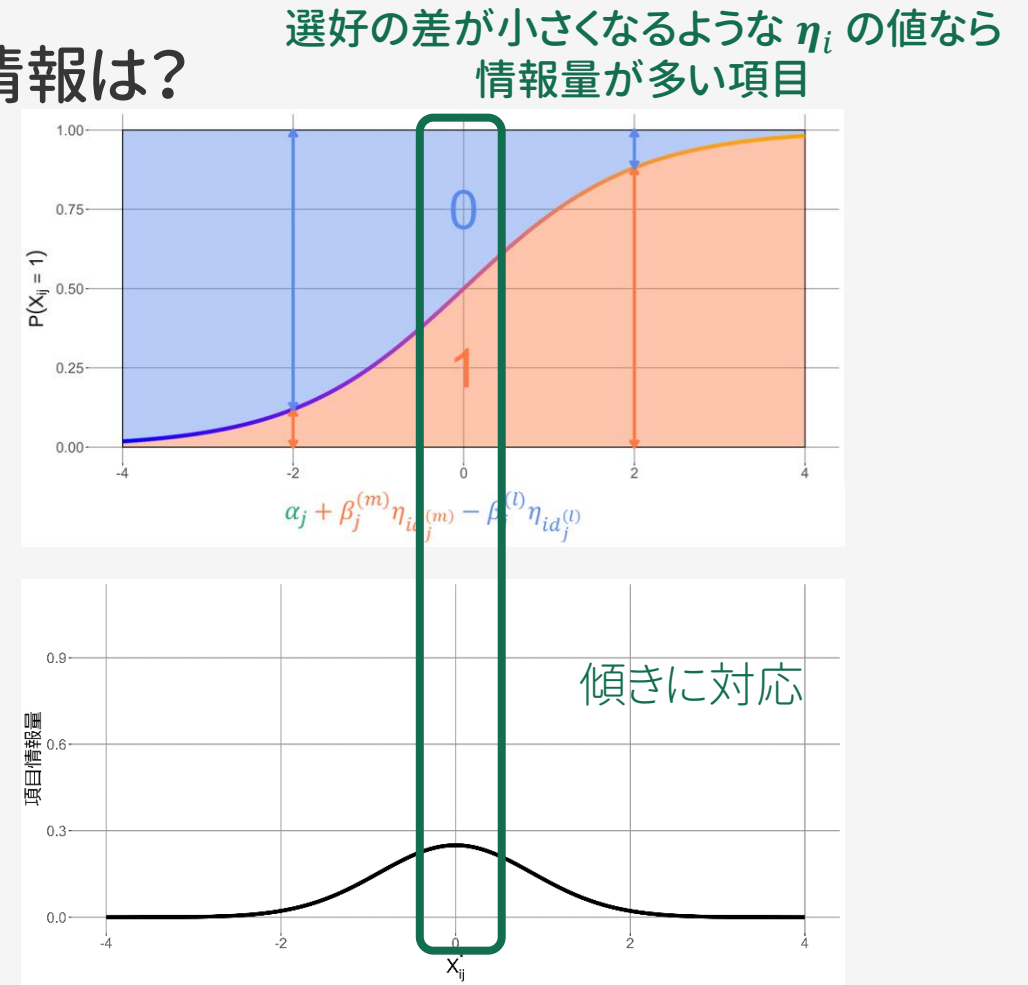
① 二値の場合



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



- 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

Q.どっちがどの程度当てはまる？

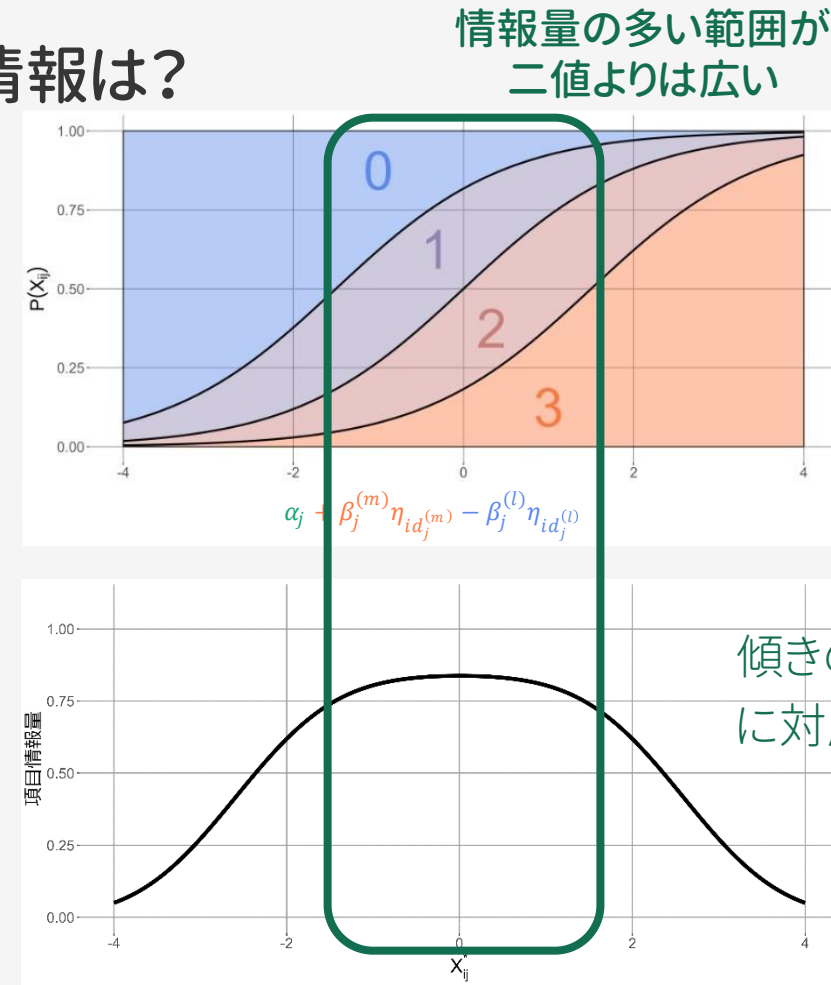
明るい	0	1	2	3	まじめ
	とても	やや	やや	とても	

$j^{(l)}$: 外向性 誠実性: $j^{(m)}$

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

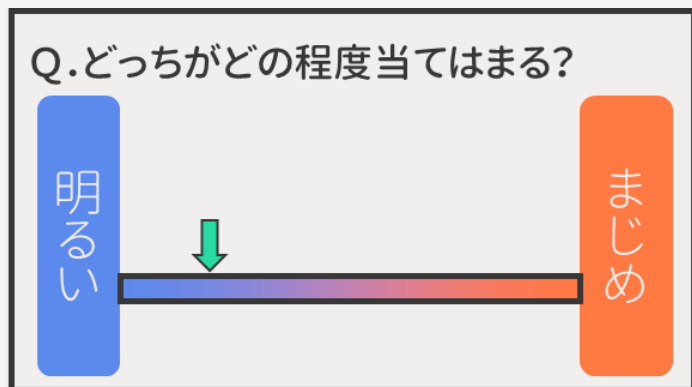


- 測定の目的は回答者の特性値を精度良く推定すること

η_i

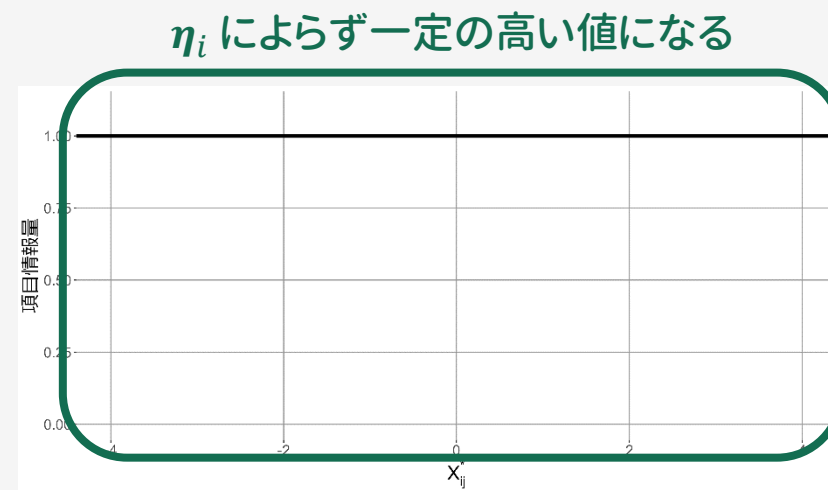
- 項目への回答が特性値推定にもたらす情報は？

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



- 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

- 項目情報量=フィッシャー情報量

▶ 対数尤度の η_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

$$I_j(\eta) = -E \left[\frac{\partial^2}{\partial \eta \partial \eta^\top} \ln P(X_{ij} | \eta_i) \right]$$

- テスト全体での情報量は単純に和を取れば良い

$$I(\eta) = \sum I_j(\eta)$$

- 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 $I(\eta)^{-1}$ ➡

項目情報量が多いほうが嬉しい

■ 具体的な項目情報量の式は

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

= 因子負荷量 × 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)

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

- 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 $I_j(\eta) = \beta_j^2$

Q.どっちがどの程度当てはまる?

明るい	3	2	1	0	まじめ
	とても	やや	やや	とても	

$j^{(l)}$: 外向性 誠実性: $j^{(m)}$

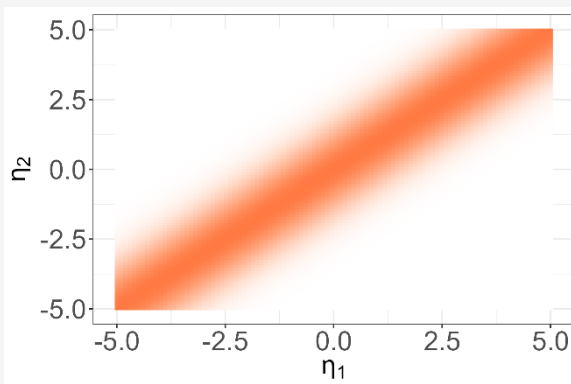
■ 段階反応モデルにおける項目情報量

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 $\mathbf{I}_j(\boldsymbol{\eta}) = \boldsymbol{\beta}_j^T \boldsymbol{\beta}_j \sum_{c=1}^{C-1} \frac{[p'(X_{ij} \geq c | \eta_i) - p'(X_{ij} \geq c - 1 | \eta_i)]^2}{P(X_{ij} \geq c | \eta_i) - P(X_{ij} \geq c - 1 | \eta_i)}$



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

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

■ 回答フォーマット ▶ 全てクリック1回で回答

二択

どちらの方が、より強く当てはまりますか

静かな方がいい

悩みだすと止まらない

強く当てはまる

強く当てはまる

多肢選択

どちらの方が、どの程度強く当てはまりますか

静かな方がいい

悩みだすと止まらない

非常に

とても

やや

やや

とても

非常に

リッカート

以下の文は、あなたにどの程度当てはまりますか

静かな方がいい

全く
当てはまらない

当てはまらない

あまり
当てはまらない

やや
当てはまる

当てはまる

非常に
当てはまる

スライダー

どちらの方が、どの程度強く当てはまりますか

静かな方がいい

悩みだすと止まらない

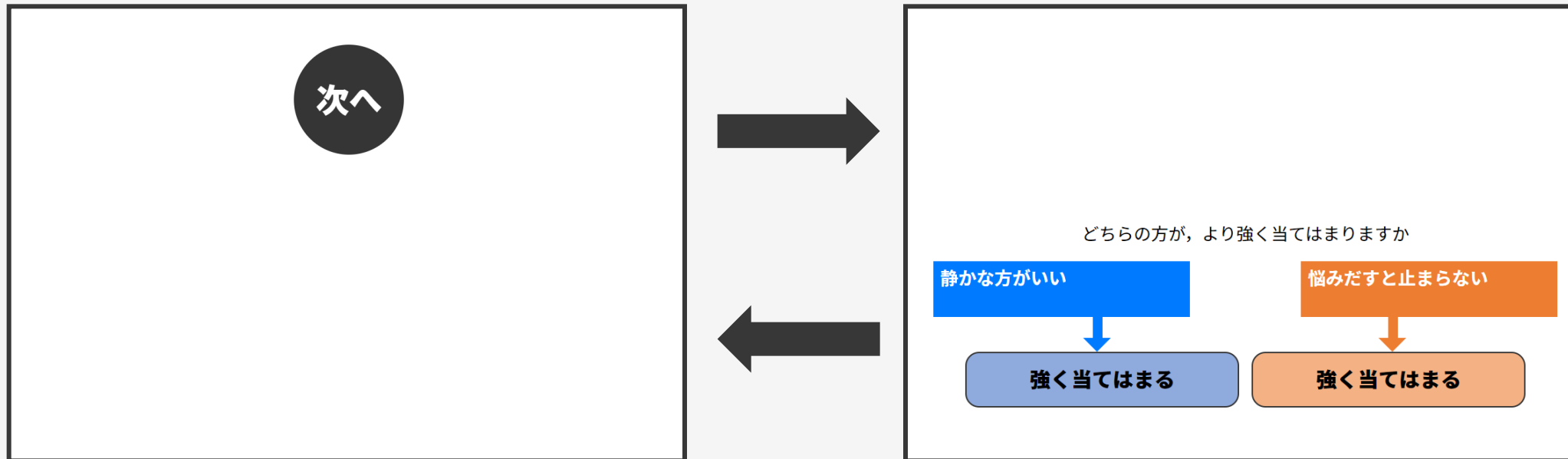
非常に

どちらともいえない

非常に

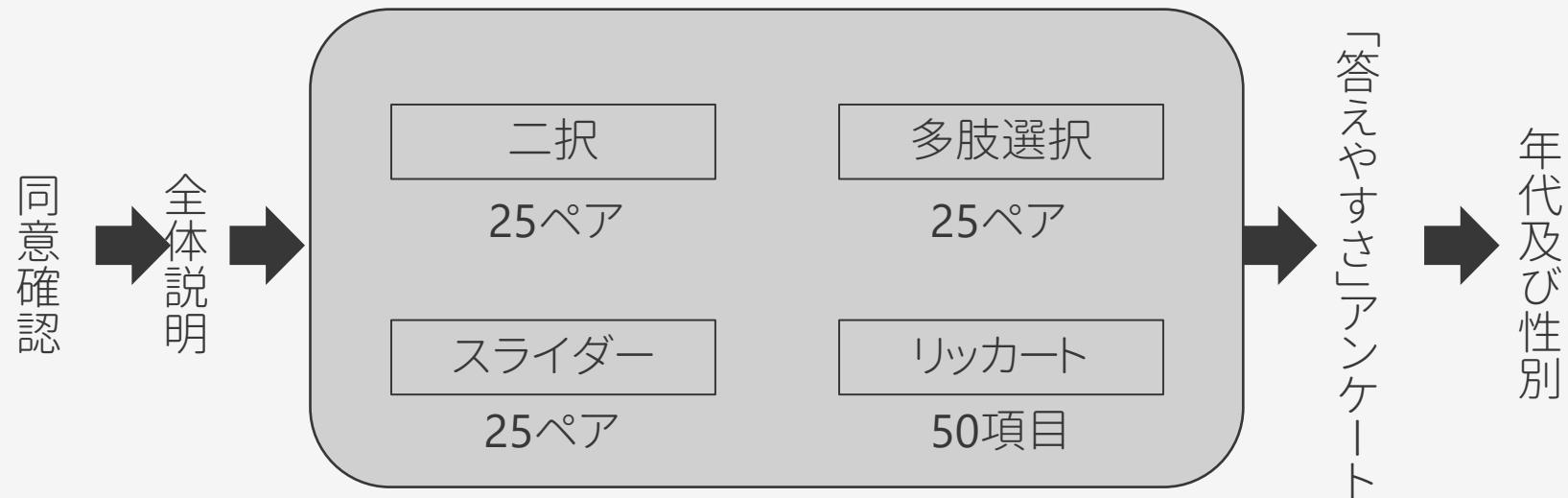
■ 各項目の間に注視点のようなものを用意

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



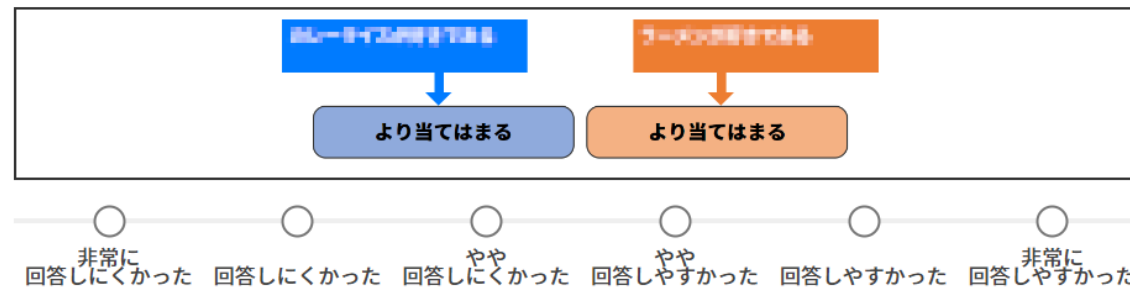
■ 実施順はランダム

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

最後に、本実験で使用した4種類の回答形式について、どの程度回答しやすかったかを6段階で評価してください。



詳細な説明は省略しますが、それぞれのフォーマットに対して分析するための統計モデルがそれぞれ存在する,ということだけ伝わればOKです。

1 二択・多肢選択

正規累積
TIRTモデル

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

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

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

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

$$X_{ij}^* = \log \left(\frac{X_{ij}}{1 - X_{ij}} \right) \leftarrow \text{適当な変換}$$

Samejima, 1973

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

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

テスト情報量は総和+事前分布の情報量

$$I(\eta) = \sum I_j(\eta) + I_\Sigma$$

1 二択・多肢選択

ロジスティック
TIRTモデル

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

2 リッカート

ロジスティック
段階反応モデル

$$I_j(\eta) = \beta_j^2 \sum_{c=1}^{C-1} \frac{[p'(X_{ij} \geq c|\eta_i) - p'(X_{ij} \geq c-1|\eta_i)]^2}{P(X_{ij} \geq c|\eta_i) - P(X_{ij} \geq c-1|\eta_i)} \rightarrow I_j(\eta) = \begin{bmatrix} I_j(\eta) & 0 \\ 0 & 0 \end{bmatrix}$$

3 スライダー

連続反応モデル

$$I_j(\eta) = \beta_j^\top \beta_j$$

Samejima, 1973

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因子全てまとめたの情報量

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 $\frac{\det(\mathbf{I}(\boldsymbol{\eta}))}{T_{\text{sec}}}$

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_{sec} 倍する $\left[\frac{1}{T_{\text{sec}}} \mathbf{I}(\boldsymbol{\eta}) \right]^{-1} = T_{\text{sec}} \mathbf{I}(\boldsymbol{\eta})^{-1}$

因子単位

回答しやすさ

対応のある一要因分散分析

■ 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

$$\alpha_j \sim \text{Normal}(0,5)$$

$$\beta_j \sim \text{student_t}(4,0,2.5)$$

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

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

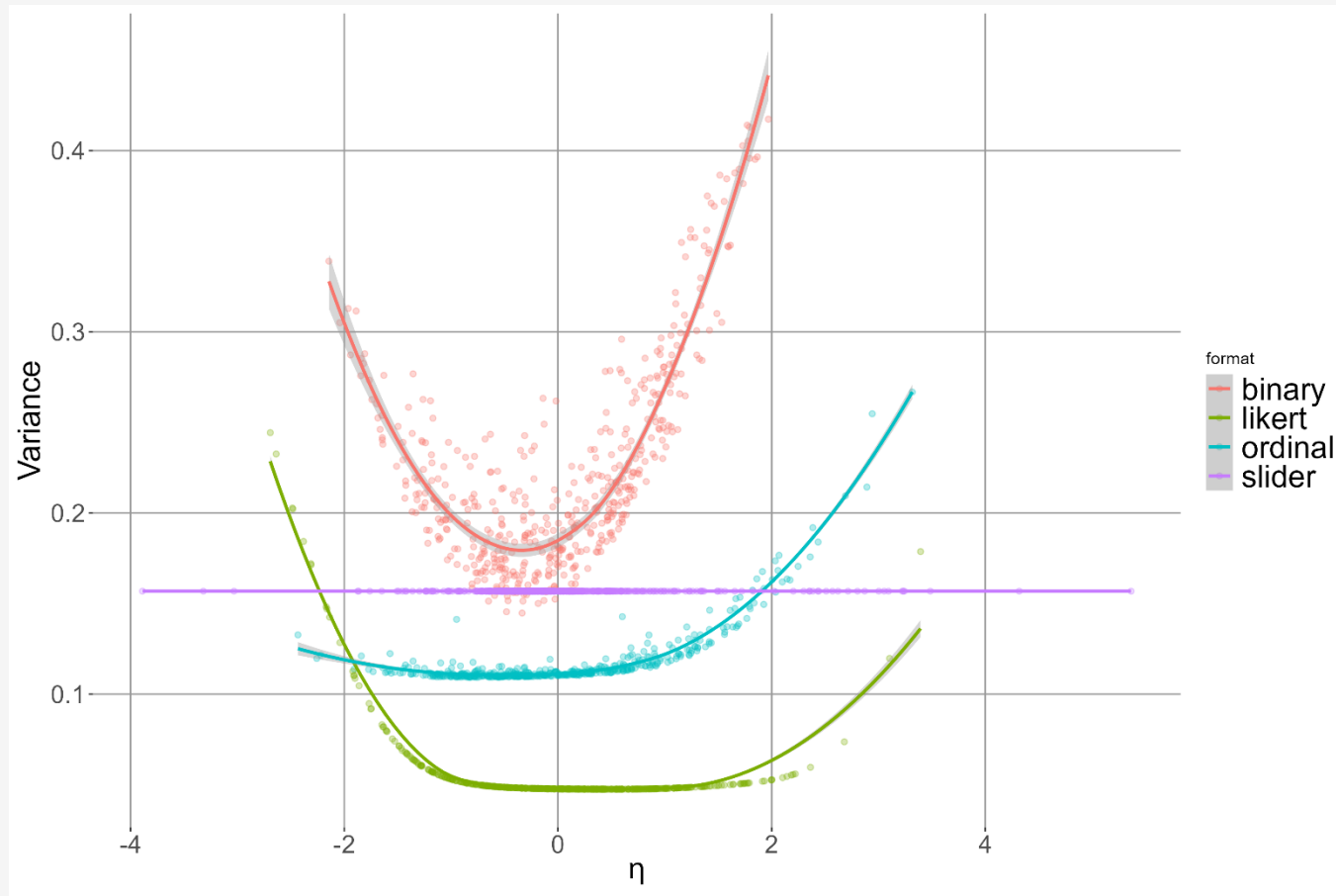
$$\gamma_j \sim \text{Normal}(0,5)$$

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

■ MCMCの設定

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

■ 分散共分散行列 $I_j(\eta)^{-1}$ 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

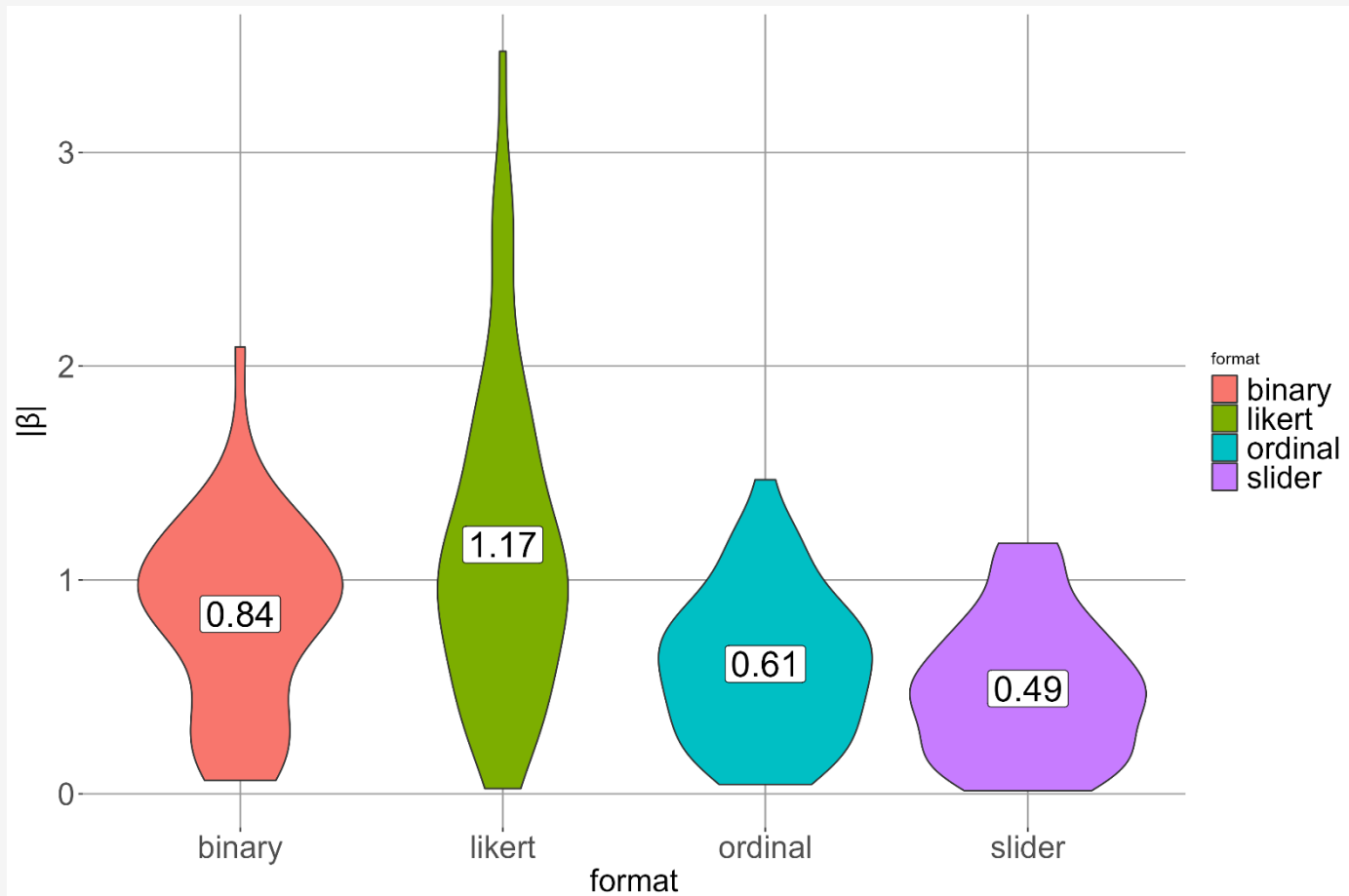
▶ 項目辺り情報量を比べても同じ結果

- 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

■ 比較型では因子負荷が低くなる

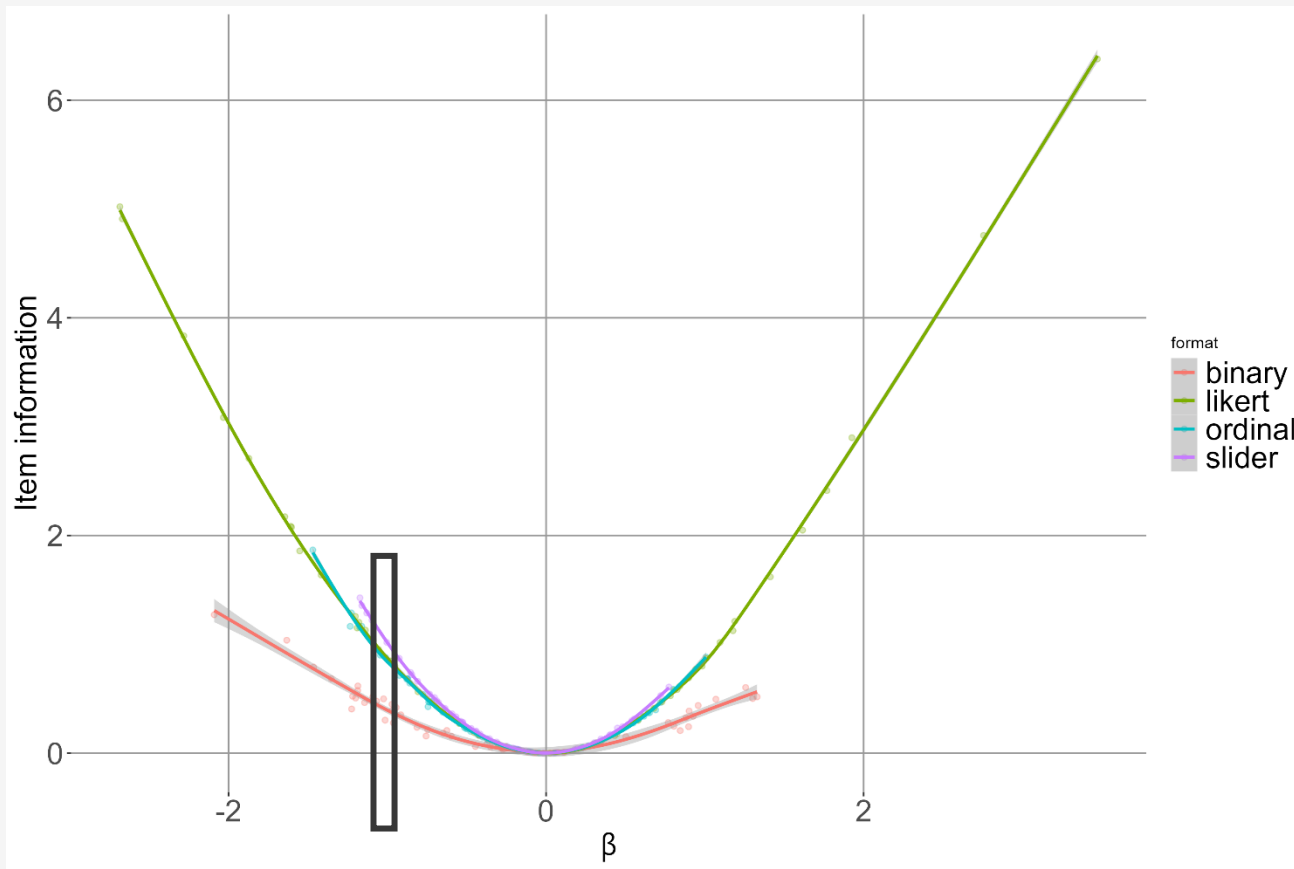
▲ 回答が「もう一方の文」によっても変わるため



- 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

■ 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

■ 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
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
リッカート	128.67	123.86	49.81

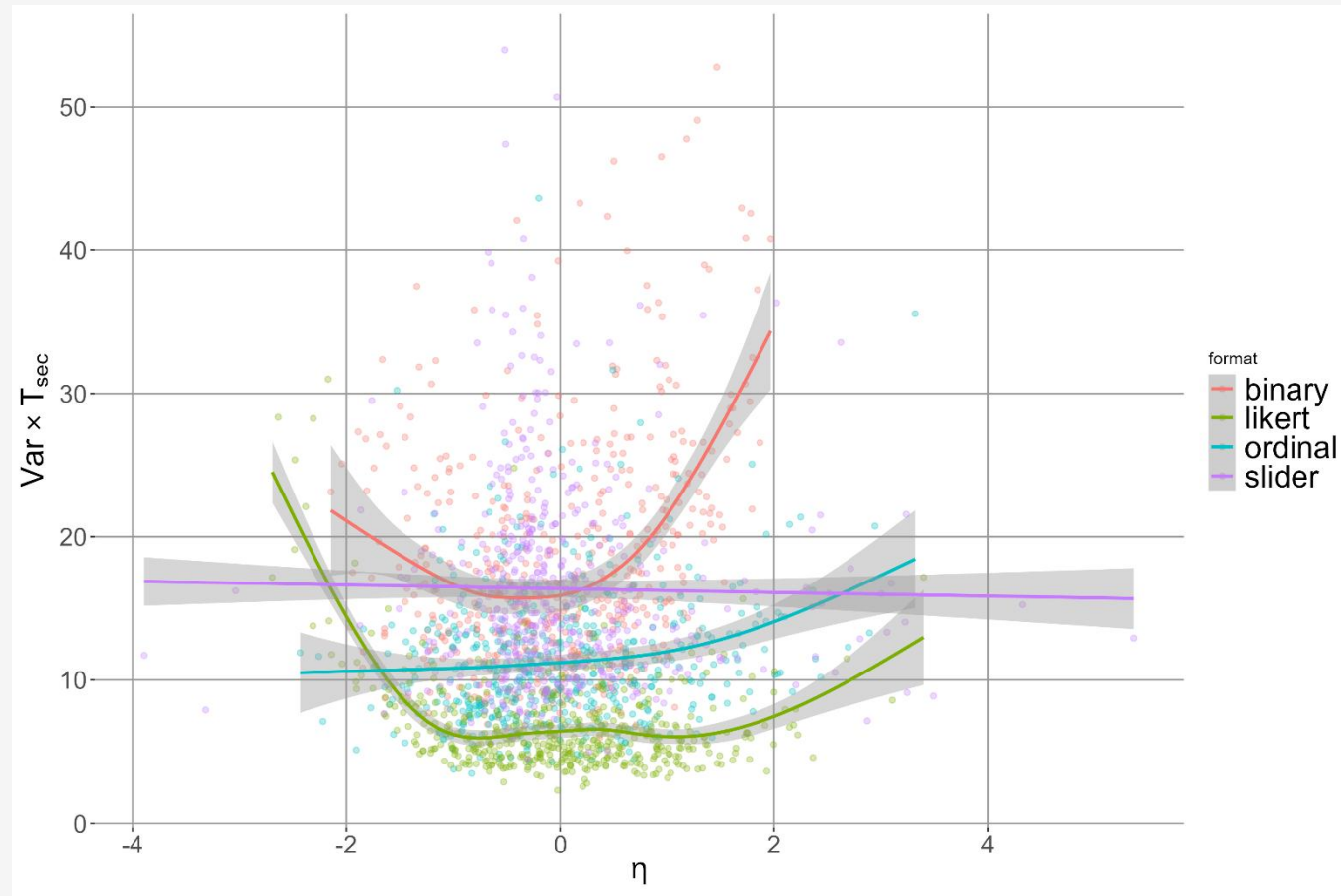
25問

50問



- 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

■ 分散共分散行列 $I_j(\eta)^{-1}$ 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

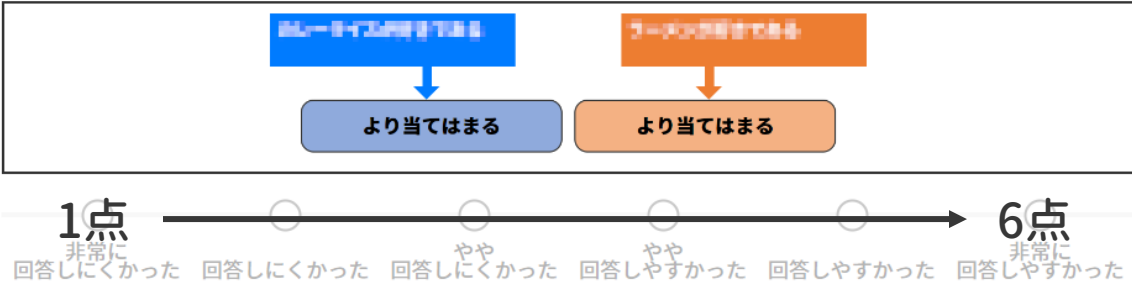
単純集計

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		

最後に、本実験で使用した4種類の回答形式について、どの程度回答しやすかったかを6段階で評価してください。



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

■ 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?

■ 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...

比較型項目における 項目要素の長さに関する検討

分寺 杏介¹ 杉山 剛² 岡田 謙介³

事前登録済み(embargo中)
ただし今回は速報的に報告します。

¹神戸大学 経営学研究科

²株式会社リクルートマネジメントソリューションズ

³東京大学 教育学研究科



¹bunji@bear.kobe-u.ac.jp

■ 特に項目要素が長いと一層答えるのが難しくなりそう

どちらのほうが,どの程度より強くあなたに当てはまりますか？

もし自分が嫌いな人に何かをしてほしいときには,そのために,その人に対して愛想よく行動するだろう。	たとえそうすればうまくいくと思っても,仕事の上で昇進するためにお世辞を言ったりしようとは思わない。				
					
Much more	More	Slightly more	Slightly more	More	Much more

■ 実際に,ペアを変えて2回連続して回答させてみたところ

モデルで推定された特性値の相関が0.4などになってしまった

▶ 人類に,短文の比較は難しすぎるのか？

■ 項目要素の長さを変えるとどうなるか？

回答の安定性はどう変化するか？ ► 再テスト信頼性による検討

回答しやすさはどう変化するか？ ► 主観的に回答してもらう

クラウドソーシング (Prolific) を利用

■ 条件と使用した項目 (4条件にランダムに割当)

【項目要素】 【フォーマット】		HEXACO (Ashton & Lee, 2009)	HEXACO形容詞版 (Romano et al., 2023)
		短文	形容詞
60項目をランダムに出題	リッカート	リッカート・短文 ($n = 229$)	リッカート・形容詞 ($n = 233$)
事前に30ペアを作成して ランダムに出題	一対比較	一対比較・短文 ($n = 239$)	一対比較・形容詞 ($n = 207$)

事前のペア設計では

- ・因子ペアの出現回数 (15ペアが2回ずつ登場するように)
- ・項目要素の符号の向きの組み合わせ (異なる向きのペアが6以上出現するように)

という制約のもとで、社会的望ましさ (予備調査で収集) の差が最小になるように遺伝的アルゴリズムで探索しました

	HEXACO (Ashton & Lee, 2009)	HEXACO形容詞版 (Romano et al., 2023)
	短文	形容詞
リッカート	How well does the following describe you? Having a lot of money is not especially important to me. Strongly inaccurateInaccurateSlightly inaccurateSlightly accurateAccurateStrongly accurate	How well does the following describe you? innovative Strongly inaccurateInaccurateSlightly inaccurateSlightly accurateAccurateStrongly accurate
一対比較	Which statement describes you better, and to what extent? When it comes to physical danger, I am very fearful. I would get a lot of pleasure from owning expensive luxury goods. Much moreMoreSlightly moreSlightly moreMoreMuch more	Which statement describes you better, and to what extent? classicaltraditional Much moreMoreSlightly moreSlightly moreMoreMuch more

■ 調査の流れ

2回の参加を求める調査

項目の提示順(および一対比較の場合は表示の左右)は毎回ランダムイズ

▲ 1回目の回答完了後,24-48時間空けて全く同じ項目セットに回答

1回目の回答終了後,「答えやすさ」を7件法で回答してもらった

■ 分析

分析の本番では,IRTモデルでベイズ推定した特性値の相関パラメータの事後分布で評価する予定です。

1. 2回の回答の単純な相関(再テスト信頼性)

個人ごとに相関を計算して,条件ごとに平均値を求める ► 単純比較

2. 答えやすさの平均値 ► 単純比較

~~3. 回答時間 ► 対数変換して単純比較~~

■ 平均値(SD)と単純比較の結果

相関係数の平均値の比較の是非はさておき……

	短文	(多重比較)	形容詞
リッカート	0.830 (SD = 0.123)	$p = 0.061$	0.864 (SD = 0.100)
(多重比較) Tukey	$p < .001$		$p < .001$
一対比較	0.677 (SD = 0.177)	$p = 0.654$	0.693 (SD = 0.169)

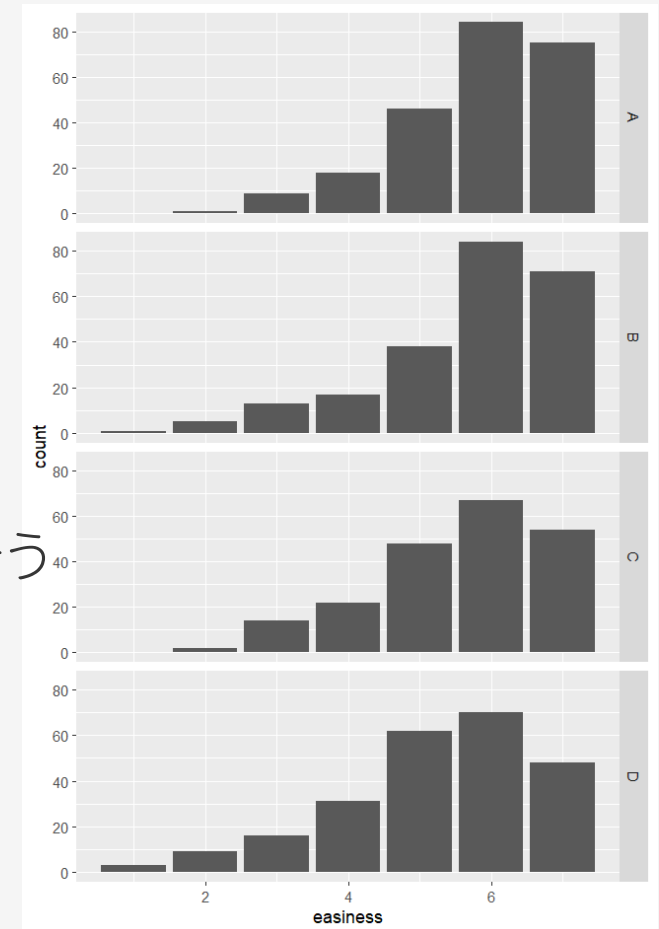
- 一対比較では明らかに測定の安定性が低下している
- 短文か形容詞かについてはあまり影響はなさそう
とはいえやはり短文のほうが少し相関係数は低下している

■ 平均値(SD)と単純比較の結果

相関係数の平均値の比較の是非はさておき……

	短文	(多重比較)	形容詞
リッカート	5.72 (SD = 1.29)	$p = 0.734$	5.84 (SD = 1.11)
(多重比較) Tukey	$p < .001$	$p < .001$	$p = 0.133$
一対比較	5.27 (SD = 1.40)	$p = 0.052$	5.57 (SD = 1.23)

- 短文の一対比較はリッカートよりもやや答えにくいと感じられそう
ただ全体的に答えやすいと感じられている?
- 短文か形容詞かはさほど影響なさそう
どちらかというに一対比較のほうが差はありそうだが



■ リッカートと一対比較では明らかに回答の安定性が変わる

一対比較は回答決定に影響する要素が単純に多いので当然ではある
だからこそ、一対比較の尺度構成では「答えやすさ」がさらに重要となる

■ 項目要素の長さの違いの影響はあまり大きくなさそう

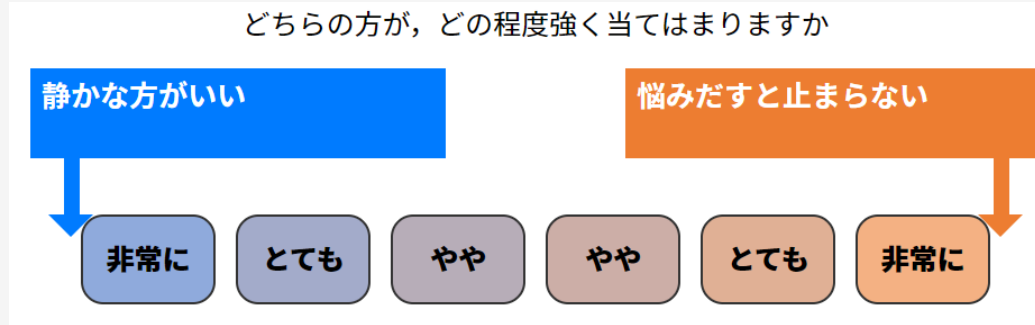
見方を変えると、無いわけでもなさそう

ただ2つくらいならある程度(やる気を出せば)対応はできそう

■ 一対比較は基本的に「難しい」回答方法

もちろん妥当性を損ねない範囲で！

▶ 様々な手を尽くしてできるだけ「答えやすい」ようにしてあげることが重要
それは「ペアの構成の方法」と「フォーマット」によって左右されそう



【研究1で使用したフォーマット】

色をつけたほうが「どちらがどの程度」感が直感的に伝わる？
一方で「色に引っ張られる」ならば良くないかもしれない

■ 実は一対比較型測定に限った話でもないと思っています。

選択肢の配置, フォント, サイズのバランス, 色, 要素の大きさ etc...

項目さえ同じならば「同じ尺度で測定した」と言って良いのか？その根拠は？