

全国規模の学力調査における重複テスト分冊法の展開可能性について（東北大学）の概要

分析の視点

わが国において、国際的な学力調査（PISA等）で用いられている調査手法（重複テスト分冊法^{※1}および項目反応理論^{※2}）を適用し、学力の経年比較や国語等の記述式問題による調査・分析を行うことは可能か？

分析方法：過年度の調査研究によって取得した重複テスト分冊法の実施ノウハウを基盤として、

- 1) 数学においては昨年度との学力の経年比較
- 2) 国語においては数学以外の教科および記述式問題に対する重複テスト分冊法の適用可能性の検証とPISA型学力を問う問題も含んだ複数の領域・内容にわたる学力の測定
- 3) 上記2)のような従来の定義よりも広範囲にわたる国語学力と数学学力との関係の把握

※1 重複テスト分冊法については柴山ほか(2011; http://www.mext.go.jp/component/b_menu/shingi/giji/_icsFiles/afieldfile/2011/10/25/1312404_01.pdf)を参照。

※2 項目反応理論とは、児童・生徒の学力と問題項目の難易度などの特性とを切り離して扱うことによって、異なる時点、冊子、児童・生徒集団等で実施した場合であっても、その結果を相互に比較できるようにするために使われる技術体系のことである。

分析結果

- ① 重複テスト分冊法を利用した学力の経年比較を実施し、複数年にわたって学力を相互に比較するために必要な技術（等化）の基本的なノウハウを獲得できた。

【技術的要素に関して今回の調査で明らかになった点】

- ・ 年度の異なるテスト間に共通の項目群を配置して経年比較を行う共通項目デザインの下で、5種類の計算方法により比較を行った。その結果、どの方法でも大きな違いはなく、計算作業の容易さ等から項目固定法^{※3}を採用した。
- ・ 項目反応理論を用いる場合と用いない場合で2年間の学力比較を行った。項目反応理論を用いると、問題の難易度の差によらず、学力が比較可能であることが確認できた。



年度間比較をするためのテストデザイン
(網掛け部分が共通項目)

※3 すでに得られているテストデータ（この場合は2010年の数学テスト）から推定した項目（図1斜線部分）に関する情報を所与のものとして利用し、テスト同士（2010年テストと2011年テスト）を比較可能な形にする方法

- ② 1つの問題文と関連する複数の設問から成る問題形式（大問形式）や、正誤ではなく部分点等の段階的な評価をすることの多い記述式問題を含む科目の例として国語を取り上げ、様々な問題形式や評価方法に対し、重複テスト分冊法や項目反応理論を適用するための方法を確認した。

【技術的要素に関して今回の調査で明らかになった点】

- ・ 各ブロックを大問形式とし、ブロック単位で分析することによって、大問形式の問題であっても、実用場面に耐えうる形で項目反応理論が適用可能であることを明らかにした。
- ・ 部分点等の段階評価を行う問題は、部分点をそのまま扱えるモデルによって分析を行った方が、部分点を与えずに正誤データ（2値データ）として扱うモデルを適用した場合と比較して、学力の低いところから高いところまで幅広い範囲の学力を扱うことが可能であることが示された。

- ③ 重複テスト分冊法の採用により、従来よりも幅広い領域の学力についての測定が可能となり、その結果、国語学力と数学学力との関係について、詳細な分析を実施することができた。数学学力は、国語学力の中でも特に説明的文章との関連が強いことが明らかになった。また、無答率については、説明的文章の選択問題の正答率が高い場合は、記述問題の無答率が低く、説明的文章の選択問題の正答率が低くても文学的文章の選択問題の正答率が高い場合は、無答率が低くなっている。

今後の展開として、重複テスト分冊法は、複数の領域・内容にわたる学力の測定が可能という利点を生かし、教科間の関連性ととどまらず、教育社会学的な様々な指標と組み合わせることによって、学力の形成要因により深く接近するための基礎情報をもたらすものであることが示唆された。

平成23年度文部科学省企画公募研究
学力調査を活用した専門的な課題分析に関する調査研究
全国的な学力調査の調査手法における技術的課題に関する調査研究

全国規模の学力調査における 重複テスト分冊法の 展開可能性について

東北大学大学院教育学研究科

柴山 直

全国的な学力調査に関する専門家会議(第4回)
平成24年7月4日水曜日10時~12時
於 文部科学省東館3F1特別会議室

項目反応理論モデルと重複テスト分冊法

(平成22年度調査研究成果)

- 項目反応理論(IRT)モデル
 - 難易度などの項目の性質 (項目特性)
 - 対象とする児童・生徒の学力 (学力特性値: 尺度値: θ)
- 重複テスト分冊法

例: 1分冊 = 10項目 × 3ブロック = 30項目

		ブロック (複数の項目が排他的に含まれている単位)					
分冊 (同じ学校・クラス であっても 児童・生徒 にランダム に付与)	分冊1	ブロック1	ブロック2	ブロック3			
	分冊2		ブロック2	ブロック3	ブロック4		
	分冊3			ブロック3	ブロック4	ブロック5	
	分冊4				ブロック4	ブロック5	ブロック6
	分冊5	ブロック1				ブロック5	ブロック6
	分冊6	ブロック1	ブロック2				ブロック6

平成23年度の目的: 展開可能性を探る

- 1) 数学においては平成22年度データとの比較
 - 学習指導要領改訂等による問題の差し替え
 - 学力の経年変化の追跡
- 2) 「国語」においては
 - PISA型問題:リーディング・リテラシー
 - 多値IRTモデルの適用
 - 記述式問題等への適用
- 3) 学力と外的指標との相関分析への展開
 - ～ 外的指標: 就学援助率、Socio Economic Statusなど
 - ～ 学力の形成要因を探る
 - ～ 相関分析の例示:
「国語」と外的指標としての数学の関係の把握

3

実施概要

- 対象学年 中学3年生
- 対象人数 1954名
- 実施校 宮城県内の5自治体の学校(20校)
- 実施時期 平成23年11月10日～29日
- 出題教科 数学・国語
- 実施時間 2授業時間(数学・国語各1時間)
- 出題範囲 学習指導要領の内容構成に準拠
国語についてはPISA型のリーディング・リテラシー問題が中心
- 分冊数 8分冊

4

分冊設計の際のブロック・デザイン

数学 1分冊 = 4項目 × 8ブロック = 32項目

		ブロック															
		B1	B2	B3	B4	B5	B6	B7	B8	B9	B10	B11	B12	B13	B14	B15	B16
分冊	S1																
	S2	○															
	S3	○	○														
	S4	○		○													
	S5				○												
	S6	○		○		○											
	S7	○	○	○													
	S8	○	○	○	○												

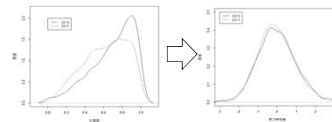
国語: 1分冊 = 4項目(記述問題含む) × 3ブロック = 12項目

		ブロック												
		B1	B2	B3	B4	B5	B6	B7	B8	B9	B10	B11	B12	B12B
分冊	K1	○												
	K2		○					○				○		
	K3			○				○					○	
	K4				○		○							○
	K4B													○
	K5	○										○		
	K6		○				○				○			
	...K7...			○				○					○	
K7B													○	
K8				○				○				○		

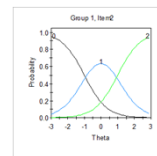
5

結果の概要

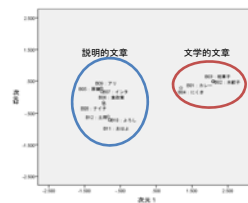
1) **テスト等化**により、項目群の難易度の変化に関わらず **二つの集団分布が比較可能**



2) **多値IRTモデル**の導入により、**記述問題**への対応が可能



3) **重複テスト分冊法**により、幅広い領域が調査できることから **様々な分析**が可能



6

1)異なる2つの集団の比較可能性

- 対象集団の変更
 - 平成22年度データ
 - 新潟市
 - 中学校 22校
 - 生徒数 2394名
 - 平成23年度データ
 - 宮城県内の5つの自治体
 - 中学校 20校
 - 生徒数 1954名
- 項目側の変更
 - 学習指導要領の改訂あり
 - 平成22年度のデータより判断して、削除する項目あり
 - 平成23年度調査のために追加する項目あり

7

数学 平成22/23年度差し替え項目一覧

56(H22) - 12(削除) + 20(新問) = 64(H23)

2010年度 項目番号	識別力	困難度	削除理由
B0-1	1.294	-1.634	昨年正答率が予想より高かったため。
B0-2	1.065	-1.813	小学校の範囲のため。
B0-7	1.611	-1.510	記号を使うこと以外は小学校の範囲のため。
B0-12	1.499	-1.269	他の設問と重なるため。
B0-14	1.144	-1.580	想定正答率より高かったため。
B1-6	0.879	-1.225	既題があったため。
B2-2	1.614	-1.164	記号を使うこと以外は小学校の範囲のため。
B2-3	0.641	-0.753	既題があったため。
B3-3	1.378	-1.597	想定正答率より高かったため。
B3-6	0.462	-0.645	識別力が低かったため。
B4-7	1.288	-1.247	既題があったため。
B5-5	1.156	-1.343	新問を入れる段階で、調整問題となった。

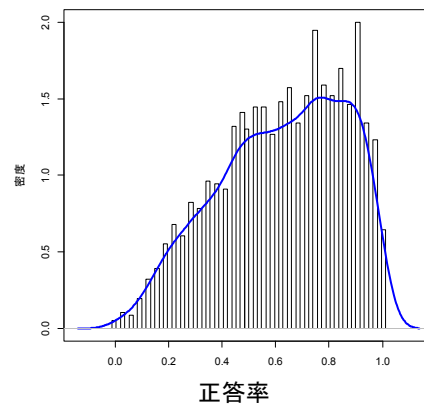
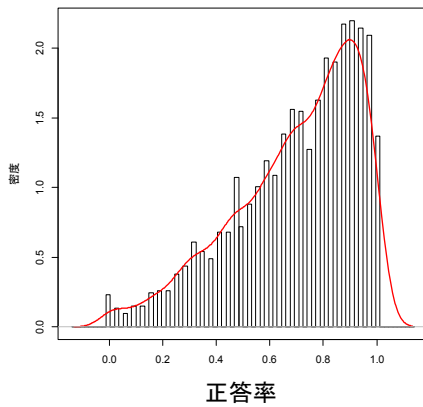
↳ H23年度の方が難しい構成となると予測

8

正答率による集団比較(数学)

平成22年度集団

平成23年度集団

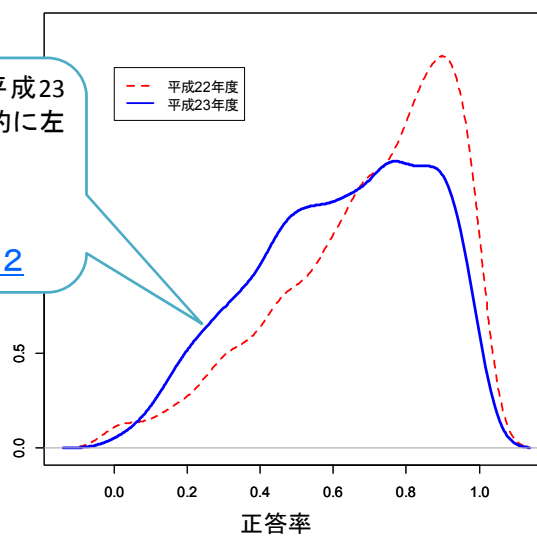


正答率による集団差の様子

難しいテストとなった平成23年度集団の方が全体的に左側に移動している

平均正答率:

0.70 → 0.62



IRT等化の基本的な考え方

- 等化のためのデータ収集デザイン

- 共通項目デザイン

- 共通受検者デザイン

- 係留テストデザイン

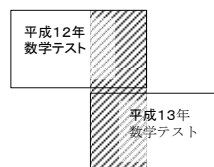
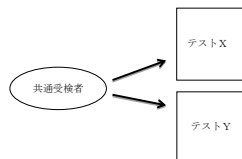
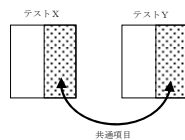
×

- 等化の計算方法

- 同時尺度調整法

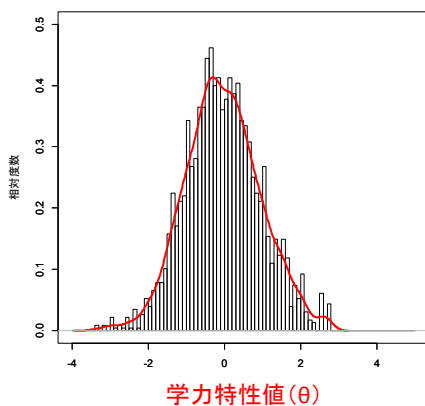
- 項目固定法

- mean / sigma法

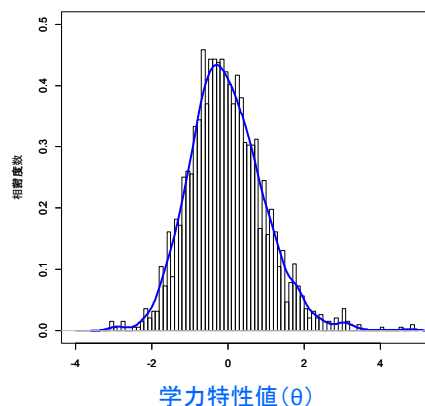


テスト等化による集団差の比較

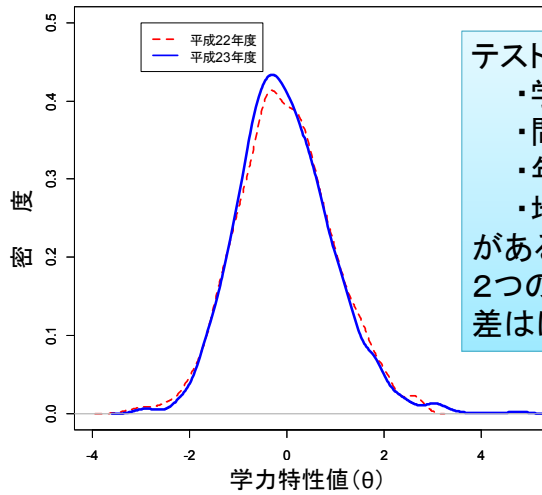
平成22年度



平成23年度



テスト等化による集団差の様子



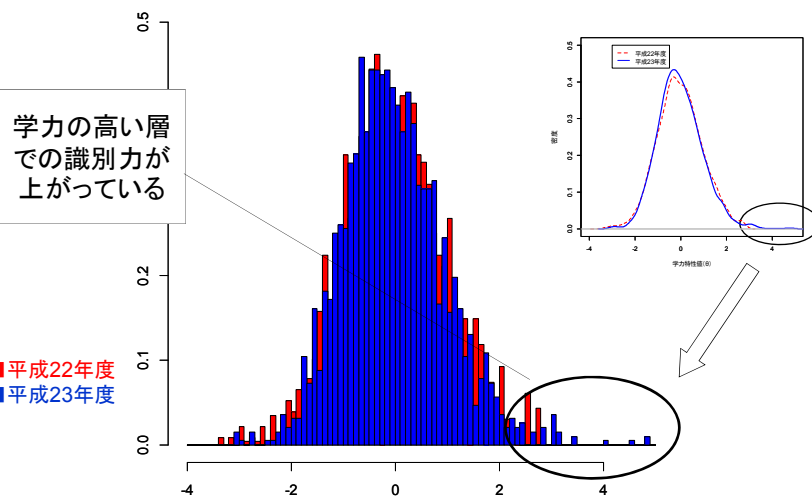
テスト等化の結果、

- ・学習指導要領の改訂
- ・問題の入れ替え(易→難)
- ・年度の違い
- ・地域の違い

があるにも関わらず、
2つの調査対象集団間に
差はほとんどないと判断できる

学力特性値		
	平成22年度	平成23年度
平均	-0.033	-0.025
標準偏差	0.985	0.978

分布の重ね合わせによる考察



学力の高い層
での識別力が
上がっている

■平成22年度
■平成23年度

2) 多値IRTモデルの導入可能性

- Samejima(1969)の段階反応モデルを採用
- 通常のIRTモデルでは正誤のみ(2値)
- 採点基準による段階的な評価(多値)
 - ⇒ ある段階に入る確率をモデル化
- モデル式(抜粋:詳細は報告書参照)

$$P(u_j = k|\theta) = P_{jk}(\theta) = P_{jk}^*(\theta) - P_{jk+1}^*(\theta)$$

$$P_{jk}^*(\theta) = \frac{1}{1 + \exp[-Da_j(\theta - b_{jk}^*)]}$$

$$P_{j0}(\theta) = P_{j0}^*(\theta) - P_{j1}^*(\theta) = 1 - P_{j1}^*(\theta)$$

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

$$b_{jk} = \frac{b_{jk}^* + b_{jk+1}^*}{2}$$

$$b_{j0} = b_{j1}^*$$

$$b_{jK-1} = b_{jK-1}^*$$

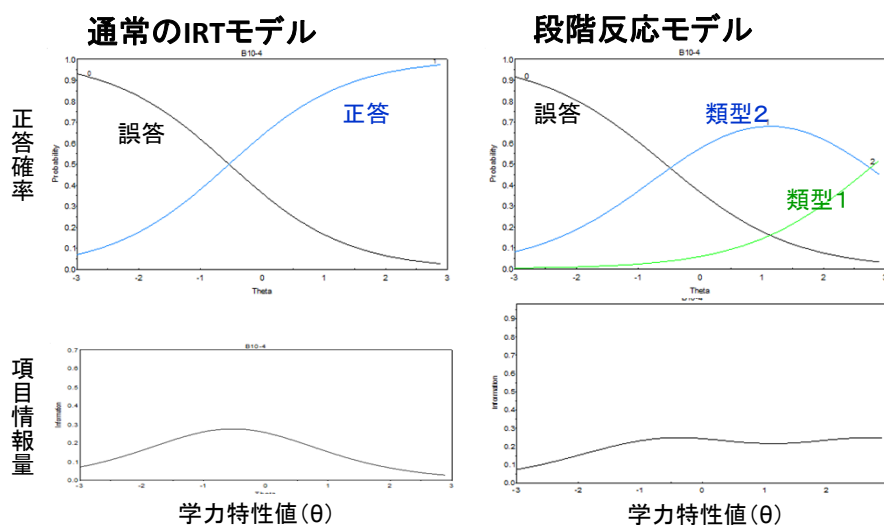
PISA調査におけるリーディング・リテラシーと本調査の関係

PISA調査(2000) 形式的側面	PISA調査:分析時における3側面	本調査研究における定義
1 情報の取り出し	情報の取り出し [PISA2009: 情報へのアクセス・取り出し]	定義1 情報の取り出し
2 幅広い一般的な 理解の形成	解釈	定義2 解釈
3 解釈の展開		
4 テキストの内容の 熟考・評価	熟考・評価	定義3 情報の編集・統合
5 テキストの形式の 熟考・評価		定義4 判断

具体例: 報告書 資料編 p.94 資料3.2
公開問題 国語(ブロック10) 問4

- 解答を満たすべき条件
 - A) 本文において「控えめな敬語」は「よろしいでしょうか」ではなく「よろしかったでしょうか」である、ということについて正しく理解して解答している。
 - B) 図のデータから必要な根拠を正しく抜き出して解答している。「変な言い方だ」と思っている人が多い。「ていねいだ」と思っている人が若い人に少ない。など)
- 正答
 - 類型 1: 条件を全て満たしているもの。
 - 類型 2: Aについて触れずに、Bを満たしているもの。
- 2値データの場合はいずれも正解とした
- 多値データの場合は
誤答 → 類型2 → 類型1 の3段階

段階反応モデルの有効性

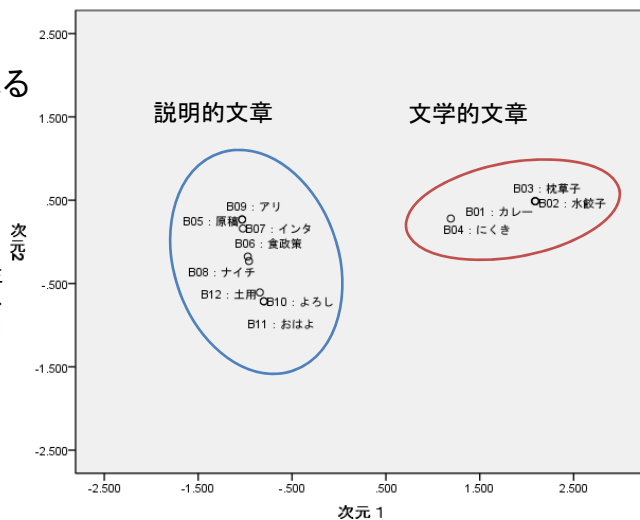


3) 学力と外的指標との相関分析への展開(例:国語と数学等)

国語の問題文の 多次元尺度法による 類似性マップ

手続き

- 1) 心理学系の9名の学生に12個の題材分を教示しソーティング法により自由に分類してもらって得た類似性行列を入力データとした。
- 2) SPSS/ALSCALにより布置を求めた。



小問ごとの数学θとの相関係数

ブロック	文章の質	項目			
		1	2	3	4
1	文学的文章(小説)	.098	.090	.173	.114
2	文学的文章(小説)	.165	.122	.188	.226
3	文学的文章(古典)	.338	.181	.098	.163
4	文学的文章(古典)	.281	.327	.298	.404
5	説明的文章(連続)	.233	.350	.235	.330
6	説明的文章(連続)	.265	.285	.293	.270
7	説明的文章(連続)	.360	.011	.305	.039
8	説明的文章(非連続)	.240	.239	.292	.214
9	説明的文章(連続)	.281	.254	.397	.386
10	説明的文章(非連続)	.222	.223	.355	.345
11	説明的文章(非連続)	.378	.224	.385	.104
12	説明的文章(非連続)	.354	.265	.427	.374

数学の学力と国語の問題との相関係数

文章の質別の本調査の国語の定義ごとに含まれる項目と
数学θとの相関係数の平均

	情報の 取り出し	解釈	情報の 編集・統合	判断
文学的文章 (小説)	0.165	0.094	0.177	0.114
文学的文章 (古典)		0.275	0.245	0.251
説明的文章 (連続型テキスト)	0.300	0.280	0.221	0.329
説明的文章 (非連続型テキスト)	0.283	0.296	0.305	0.301

21

国語の記述問題の無答率について

表8.2.6 文学的文章(選択)の高低×説明的文章(選択)の高低ごとの記述問題の
無答率及び誤答率の平均とSD

文学的文章(選択問題)	低群(n=1158)				高群(n=796)			
	低群(n=556)		高群(n=602)		低群(n=195)		高群(n=601)	
説明的文章(選択問題)	平均	SD	平均	SD	平均	SD	平均	SD
記述問題の無答率	0.26	0.37	0.08	0.22	0.13	0.26	0.05	0.19
記述問題の誤答率	0.50	0.33	0.51	0.28	0.55	0.27	0.44	0.27

無答率について

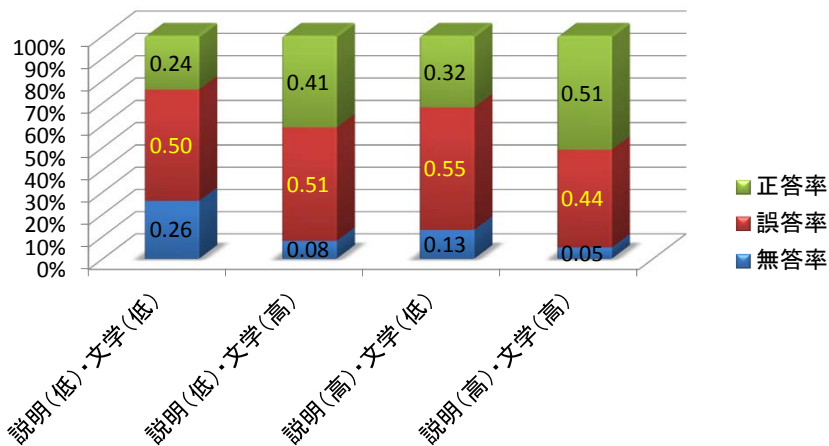
説明的文章(選択問題)低群 > 説明的文章(選択問題)高群

このうち

文学的文章(選択問題)低群 > 文学的文章(選択問題)高群

国語の記述問題の無答率について

説明的文章(選択問題)の高低群×文学的文章(選択問題)の高低群ごとの、記述問題の正答率・無答率・誤答率



まとめと今後の問題

- IRT等化の導入
～ 経年変化の追跡が可能
- 多値IRTモデルの導入
～ 記述・証明問題への適用が可能
- 重複テスト分冊法の採用
～ 幅広い領域が調査できることから
様々な分析が可能



マトリックス・サンプリングによる集団統計量の
推定(経年変化・マルチレベル分析等)