

平成 29 年度文部科学省委託研究
「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」
研究成果報告書

本体調査の個票データの匿名化に関する調査研究

平成 30 年 3 月



株式会社内田洋行
教育総合研究所

目次

1	背景・目的	1
1.1	背景	1
1.2	本研究の目的	1
1.3	各種データの位置付け	2
1.4	用語の定義	4
2	匿名データに関する検討	5
2.1	基本的な考え方	5
2.1.1	利用するレコードについて	5
2.1.2	利用するフィールドについて	5
2.1.3	サンプリング率について	6
2.2	児童生徒匿名データ	7
2.2.1	児童生徒匿名データの目的・概要	7
2.2.2	サンプリングについて	7
2.3	学校匿名データ	8
2.3.1	学校匿名データの目的・概要	8
2.3.2	平均正答率の丸め処理について	8
2.3.3	サンプリングについて	9
3	パブリックユースデータ（疑似データ）に関する検討	12
3.1	パブリックユースデータの目的・概要	12
3.2	保持する統計量について	12
3.3	データの生成	12
4	今後の展望	14
4.1	児童生徒データと学校データの紐付け	14
4.2	児童生徒データ，学校データの正規化とデータベース化	15

4.3	パブリックユースデータのバリエーション	16
	資料.....	17
	有識者委員会開催記録.....	17
	匿名データレイアウト	19
	調査研究体制	23

1 背景・目的

1.1 背景

平成 19 年より全国学力・学習状況調査（以下、本調査）が行われるようになって 10 年以上が経過しており、さらなる分析の充実を求める声は大きい¹。現在、本調査データを大学等の研究機関に条件付きで公開し、研究者独自の視点に基づく調査研究を促進することで、教育施策や指導改善に役立てるための検討が進んでいる。平成 29 年 3 月には、文部科学省によるガイドライン「全国学力・学習状況調査個票データ等の貸与・公表について」²（以下、「ガイドライン」）が公開された。平成 30 年度中より、本調査データの研究者への公開が開始される予定である。

一方で、本調査で扱われるデータは児童生徒や学校の学力という機微なデータを含んでおり、公開にあたっては十分な匿名化処理を行い、学校や個人が特定されないような留意が必要である。そこで、本研究では、学力調査データの匿名化に関する調査研究を行う。

1.2 本研究の目的

本調査研究では、「ガイドライン」に記述がある「パブリックユースデータ（疑似データ）」及び「匿名データ」に関する調査分析と、実際のデータ作成、それを踏まえた今後の可能性検討までを実施する。

実施にあたっては、過年度の「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」や、各都道府県において本調査の分析に携わったことのある有識者及び、匿名化処理に知見のある有識者を招聘して有識者委員会を設置し、各テーマについて検討を行った。

- 1) パブリックユースデータ（疑似データ）化に関する調査分析
- 2) パブリックユースデータ（疑似データ）の作成【1 か年分】

¹ 例えば、お茶の水女子大学教授の耳塚寛明氏は以下のように述べている。「学力水準と格差の変動を把握するために、こうした調査を行うことは国の責務だ。だが、今回の文部科学省の公表資料の内容には落胆を禁じ得ない。資料からは国が学校現場等に何を発信しようとしているかが伝わってこない。（学テでは）学校のカリキュラムマネジメント（時間割の管理）や児童生徒の自尊感情等についても重点的に質問しているのに、分析が薄い。早急な改善を期待する。」日本経済新聞 2017 年 8 月 28 日

² 個票データ等の貸与に係るガイドライン等

http://www.mext.go.jp/a_menu/shotou/gakuryoku-chousa/sonota/1386501.htm

- 3) 匿名データ化に関する調査分析
- 4) 匿名データの作成【10 か年分】
- 5) 今後の可能性検討

本報告書では、3) 4) の内容を2章にて、1) 2) の内容を3章にて、5) の内容を4章にて述べる。調査研究項目と章立ての順番が異なる理由は、有識者委員会での議論に基づき、文部科学省とも合意した上で、匿名データの検討を先に行い、それに合わせる形でパブリックユースデータ（疑似データ）の検討を行うという順序で調査研究を進めることとなったためである。

1.3 各種データの位置付け

個票データ及びパブリックユースデータ（疑似データ）・匿名データの位置づけについては、本研究では「ガイドライン」及び「貸与するデータの種類（イメージ）」資料の記述をもとに以下のとおり定義している。

パブリックユースデータ：

パブリックユースデータは、集計表からデータ生成した、個人情報は一切含まない疑似データである。研究者が匿名データの貸与を申請するためのデータ形式の確認・プログラム準備に利用するためのデータとする。大学生・大学院生の統計学学習のための操作演習にも利用可能。利用イメージを踏まえ、匿名データとデータレイアウトを揃えることが望ましい。教科調査の正答率の平均値と標準偏差等、一部統計値を保持することが望ましい。データ件数については、基本的な表計算ソフトウェアでも取り扱うことができるデータ件数とする。
(児童生徒ローデータ)

解答番号 解答用紙	国語A 正答数	国語B 正答数	算数A 正答数	算数B 正答数	国語A問1 解答類型	国語A問 1正誤	...	児童生徒 質問紙①	児童生徒 質問紙②	...
—										

匿名データ：

匿名データは、個票データに対して匿名化処理を行い、データ内容を限定したデータである。研究者の学会誌等論文への利用、大学院生の修士・博士論文への利用及び都道府県・市町村教育委員会の教育施策推進のための利用が可能なデータとする。学校・教育委員会・都

道府県の情報を削除する等、必要な匿名化処理を行い、サンプリングを行った上で提供するものとする。研究ニーズによっても求められるデータが異なることが考えられるが、本調査研究で作成する匿名データのデータセットは1種類とし、申請内容により個別に匿名データを作成することまでは対応しない。文部科学省に申請の上で、簡易な審査の上で利用することができるものとする。

(学校ローデータ)

学校ID	国語A平均正答数(リコーディング)	国語B平均正答数(リコーディング)	...	学校質問紙回答(全児童(生徒)数)(リコーディング)	学校質問紙回答(全教職員数)(リコーディング)	学校質問紙回答(全学級数)(リコーディング)	...
1							

(児童生徒ローデータ)

解答用紙番号	国語A正答数	国語B正答数	...	学校ID	国語A問1解答類型	国語A問1正誤	児童生徒質問紙①	児童生徒質問紙②	...	学校質問紙回答(全児童(生徒)数)(リコーディング)
—										

※児童生徒ローデータは、トップコーディング及びリコーディングの児童生徒数にあわせたデータ数を抽出し、各学校のローデータとする

個票データ：

パブリックユースデータ及び匿名データの作成元となるデータ。年度ごとに児童生徒個票データと学校個票データが作成されている。児童生徒個票データは毎年度100万件以上、学校個票データは小学校20,000件程度、中学校10,000件程度のレコード数となっており、学校・教育委員会・都道府県等の情報を含む。文部科学省に申請し、厳格な審査の上で利用することができるものとする。

(学校ローデータ)

ID	国語A平均正答数	国語B平均正答数	...	学校名	都道府県名	設置者名	本校・分校	へき地	...	学校質問紙回答(全教職員数)(実数)
1										

(児童生徒ローデータ)

解答用紙番号	国語A正答数	国語B正答数	算数A正答数	算数B正答数	国語A問1解答類型	国語A問1正誤	...	児童生徒質問紙①	児童生徒質問紙②	...
—										

本調査研究では、文部科学省による「ガイドライン」の考え方をもとに、過去に各都道府県において学力調査分析に携わった経験のある有識者及び、他省庁における調査データの匿名化に携わった経験のある有識者による委員会を組織し、個票データの匿名化のあり方について検討した。実際に作成した匿名データ・パブリックユースデータは、ここまで述べた「ガイドライン」及び「貸与するデータの種類（イメージ）」と異なる部分もあるが、匿名化にあたっての特定リスク等の課題を解決するために検討を重ねた結果、そのようになっている。次章以降で、本調査研究において作成した匿名データ・パブリックユースデータの内容について述べる。

1.4 用語の定義

レコード：

個票データの「行」のこと。1行が1人の児童生徒・1校の学校を表す。

フィールド：

個票データの「列」のこと。保持されている変数を表す。

トップコーディング・ボトムコーディング：

例えば学校規模が一定以上の学校については「201人以上」等、ある値に合わせることで、特定化を防ぐ方法。

リコーディング：

例えば実数回答項目を「1～20人」「21～40人」といった形でカテゴリー化することで、特定化を防ぐ方法。

2 匿名データに関する検討

2.1 基本的な考え方

本調査研究では、平成 19 年度から平成 29 年度まで、10 か年の匿名データを作成した。匿名データは児童生徒個票データをもとに作成する「児童生徒匿名データ」と学校個票データをもとに作成する「学校匿名データ」からなる。児童生徒及び学校で共通して削除すべきレコード・フィールドについて本節にて述べ、サンプリングの方法等、児童生徒と学校とで異なる考え方が必要となる部分について次節以降で記述する。

2.1.1 利用するレコードについて

本調査研究においては、匿名化の対象とするのは【公立学校のみ】とする。そのため、国立・私立学校に所属する児童生徒のレコードは削除することとする。

本調査で集計されるデータは【当日実施の児童生徒（調査実施日に調査を受けた児童生徒）】のみである。そこで、後日実施の児童生徒のレコードは削除する。一部教科のみ後日実施の児童生徒については、後日実施教科のデータのみ削除することとする。

また、平成 28 年度以前の本調査では、データを匿名化した上で研究者にデータ提供を行うことについて実施要領に記載されていないため、今回の匿名化にあたり、文部科学省より市町村教育委員会に対してあらためて、過年度のデータ公開に関する意向を調査している。上記調査において、【匿名データ公開可としている市町村教育委員会に属する児童生徒のレコード】のみを利用し、それ以外のレコードを削除する。なお、平成 29 年度からは匿名データの研究利用について実施要領に記載されるようになった。

2.1.2 利用するフィールドについて

利用するフィールドに関する基本方針としては、幅広い分析ニーズの可能性をかんがみ、学校及び設置者の特定化につながるフィールドのみを削除し、それ以外のフィールドについては手を加えず残すものとする。例えば児童生徒ローデータにおける学校質問紙回答は、学校に所属する児童生徒についてすべて同じ内容が入っている。あるいは、学校ローデータにおける児童生徒質問紙の肯定的回答割合のフィールドは、児童生徒質問紙の選択肢別回答割合から算出可能であり、冗長と思われるが、個票データにおいても当該フィールドが存在し、削除する必然性がないことから、そのまま利用する。

削除するフィールドについて、以下のとおりである。

「国公私区分」

「当日実施_ (教科)」

公立学校の当日実施児童生徒のみを匿名化の対象とするため、削除する。

「都道府県コード」「都道府県名」

「設置者コード」「設置者名」

「学校コード」「学校名」

以上のフィールドについては、学校の特定化及び都道府県・設置者間の序列化につながるおそれがあるため削除する。

「学校種」

「本校分校」

「所在地市町村コード」「所在地市町村名」

「へき地区分」

「学校の学級数規模」

以上のフィールドについては、学校に関するメタデータにあたり、組み合わせることで学校の特定化につながるおそれがあるため削除する。

「人数_ (教科)」(学校個票データのみ)

学校個票データに存在する、各教科を受験した児童生徒数を示すフィールドである。児童生徒数は学校ウェブサイト等にも公開されていることが多く、学校の特定化につながるおそれがあるため削除する。

「当日実施_ (教科)」

先述のとおり、当日実施データのみを匿名化の対象としており、当日実施かどうかを区別するフィールドである本項目は不要となるため削除する。

「学校質問紙回答_実数」

学校の児童生徒数及び教員数(全体・経験年数別)等に関する情報を含むフィールドであり、「人数_ (教科)」と同じく学校の特定化につながるおそれがあるため削除する。

その他、配慮事項に関するフィールド(点字問題冊子・拡大文字問題冊子の利用等)、及び、年度特有のフィールドを削除する。

2.1.3 サンプルング率について

抽出率の異なる複数パターン(1%・5%・10%)の匿名化サンプルデータについて、平

均正答率及びその信頼区間を検討し、有識者委員会での議論も踏まえ、サンプリング率については【10%】を採用することとする³。平均正答率の信頼区間を一定区間に収めるという意味では、より小さなサンプリング率でも十分な値が得られるが、本調査で扱われる変数は平均正答率の他にも設問別の解答類型や児童生徒質問紙回答までさまざまなデータがあり、研究者の多様な研究ニーズを保証するという意味で、匿名性が保たれる範囲で可能な限りデータ件数を多く確保することも重要であると考えられたためである。

なお、本調査が抽出形式で行われた平成 22 年度及び 24 年度のデータについても、一律で 10%の抽出率を適用する。そのため、これら 2 年分の匿名データについては他の年度よりもデータ件数が少なくなることとなる。

2.2 児童生徒匿名データ

2.2.1 児童生徒匿名データの目的・概要

児童生徒匿名データは、児童生徒を 1 レコードとする匿名データである。児童生徒単位の各教科の正答数・正答率と、設問別の正誤情報・解答類型、児童生徒質問紙への回答情報、及び児童生徒が所属する学校の学校質問紙回答情報が保持されたデータである。



図 1 児童生徒匿名データのレイアウト (イメージ)

2.2.2 サンプリングについて

10%抽出のランダムサンプリングとする。抽出後、解答用紙番号に意味のない連番を付与

³ サンプリング率については以下を参考にした。

永田靖 (2003) 統計ライブラリー サンプルサイズの決め方. 朝倉書店

山口拓洋 (2010) 臨床家のための臨床研究デザイン塾テキスト中級編 2 サンプルサイズの設計. 健康医療評価研究機構

する。

2.3 学校匿名データ

2.3.1 学校匿名データの目的・概要

学校匿名データは、学校を1レコードとする匿名データである。各学校に所属する児童生徒の正答数・正答率を集計した平均正答数・平均正答率、無解答数・無解答率、学校内の「四分位層」及び「四分位層別割合」（正答率を4分割したA層～D層に属する児童生徒の割合）、児童生徒質問紙の各質問への回答割合、そして学校質問紙回答情報が保持されたデータである。



図 2 学校匿名データのレイアウト (イメージ)

2.3.2 平均正答率の丸め処理について

学校の平均正答率は、本調査の中でも最も機微なデータの一つと言える。学校の平均正答率が公表されることにより、学校間の過度な競争や序列化が起こることは、本調査の主旨とは異なる。

平成 26 年度に文部科学省が実施した「全国学力・学習状況調査の結果公表に関する調査」⁴において、市町村全体の教科の平均正答率を公表している市町村が 28%、市町村立学校の教科の平均正答率までを公表している市町村は 2%と、少ないながらもたしかに存在している。数年前の結果ではあるものの、学校ウェブサイト等に公表されている情報を参照することで、匿名データの学校を特定化することができないとは言い切れない。

⁴ 平成 26 年度全国学力・学習状況調査の結果公表に関する調査結果について
http://www.mext.go.jp/a_menu/shotou/gakuryoku-chousa/detail/1353969.htm

平成26年度全国学力・学習状況調査の結果公表に関する調査結果 市町村教育委員会の状況

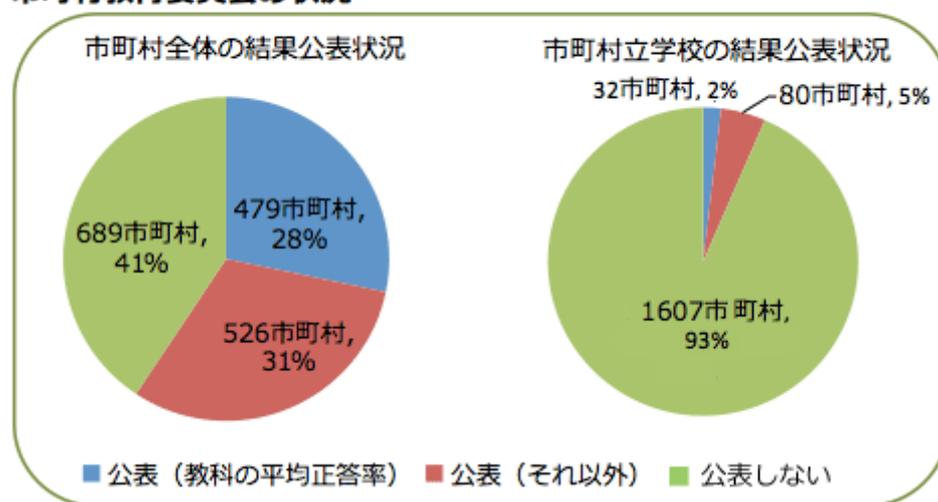


図 3 本調査の結果公表に関する状況

そこで、各種平均正答率（教科全体及び領域・観点・形式別）を整数値に丸めることとする。整数値にすることで、ある正答率の学校が1校に特定される確率が格段に下がること（例えば、平均正答率が70%である学校は他にないとは言い切れなくなること）、整数値であっても多変量解析が十分に可能な情報量を保持していること、平成29年度の本調査より、平均正答率について整数値での結果公表となっており、この流れとも軌を一にすることが理由である。

なお、平均正答数については、毎年教科の設問数が変わり解釈が異なることもあり、平均正答率と比べて公表されていないことが多いため、丸め処理は行わないこととする。

2.3.3 サンプルングについて

学校についてはランダムサンプルングとせず、地域規模及び学校規模を考慮したサンプルングを行うこととする。

【地域規模】都道府県・設置者・学校の情報を削除することにしたものの、研究ニーズとして、地域情報を含めて分析したいというニーズは一定数あることが想定される。本調査の個票データには「地域規模」というフィールドがあり、都市規模による区分（大都市・中核市・その他の市・町村）が含まれているため、この情報は残すことが望ましい。

【学校規模】極端に大規模あるいは小規模な学校は特定化のリスクが高い。大規模な学校

は全国にあまり存在しておらず、学校が特定される危険性がある。一方で、小規模な学校は1名・2名しか受験しておらず、学校の結果がそのまま個人と紐付けられるおそれがある。

まず、大規模校・小規模校について児童生徒数で一定の基準を設け、基準以上・以下の学校をデータから削除することを検討した。これにより、たしかに平均正答率の外れ値にあたる学校が少なくなることがわかったが、人口減少等により小規模校が増えていることから、これらの学校が削れることによりデータ数が2割程度減少することがわかった。そのため、学校規模による学校の一括削除は適切でないと判断した。

そこで、児童生徒数及び地域規模4区分に基づいて各学校を人数による階級に区分し、階級ごとにサンプリングする（層化抽出）という方法を取ることにした。児童生徒数については、全年度に共通して存在しておりかつ調査対象学年の児童生徒数として適切なフィールドとして「人数_児童生徒質問紙」を利用することとした⁵。具体的には、地域規模4区分と児童生徒数の階級に基づいてクロス集計表を作成し、それぞれの階級に一定数の学校が存在することを確認した上でサンプリングをすることとした。地域規模ごとにクロス集計表のセル内の学校数があらかじめ決めた基準⁶に満たなくなったところでトップコーディングを行い、例えば「201人以上」等にまとめた階級とし、そこから10%のサンプリングを行うこととした。なお、これらの基準については全年度共通とした。

⁵ 本調査は教科によって受験人数が異なる場合がある。学校の児童生徒数を最も正しく示すと思われる学校質問紙の実数回答質問「第6学年の児童数/第3学年の生徒数」は平成19年度データに存在していない。他の年度を参照すると、「人数_児童生徒質問紙」が、「人数_国語」や「人数_算数・数学」よりも、実際の児童生徒数に最も近い値が得られる。

⁶ 基準の設定にあたっては「ガイドライン」の記述を参考にした。

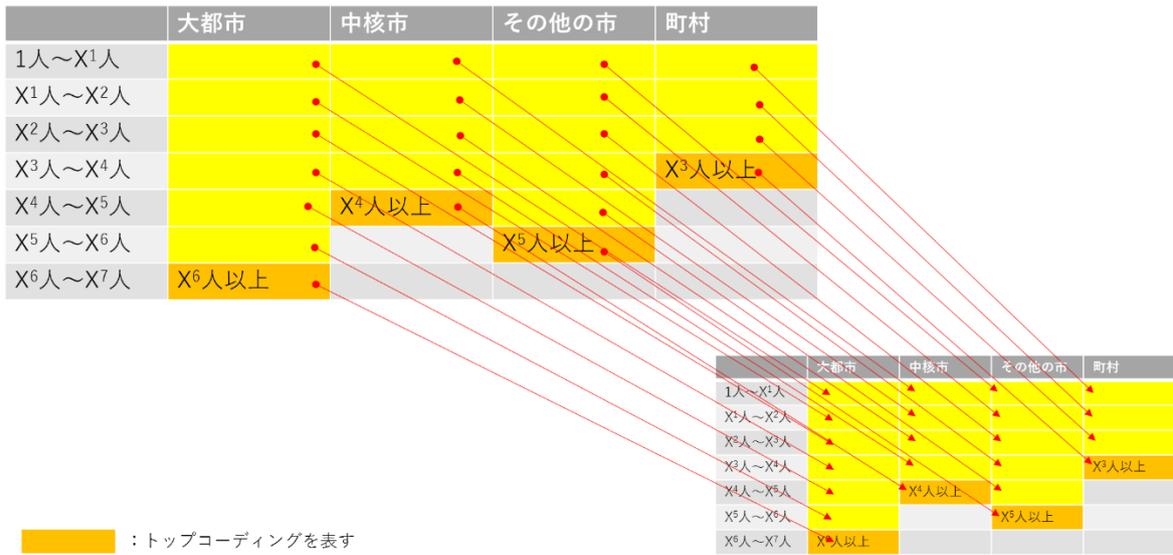


図 4 サンプリングのイメージ

3 パブリックユースデータ（疑似データ）に関する検討

3.1 パブリックユースデータの目的・概要

1章において述べたとおり、本調査研究におけるパブリックユースデータとは、児童生徒及び学校の個人情報的一切保持せず、個票データより作成した統計表を基に疑似的に生成したデータのことである。文部科学省ウェブサイト等で公開され、匿名データとデータレイアウトを同一としているため、研究者が匿名データを申請する際の事前確認及び分析プログラムの準備に利用することができる。また、大学等で統計学の操作演習等にも利用することができるが、計算した統計値は実際のデータではないため、実証研究に利用することはできないものである。

本調査研究では、平成27年度データの疑似データを作成した。27年度を対象とした理由は、全数調査で理科を含め5教科で調査実施した最新年度であったためである。

3.2 保持する統計量について

教科の平均正答率を保持することが望ましい。また、各設問の正誤・解答類型及び領域・観点・形式別の正答数・正答率等の調査結果についても、不整合がないよう合理的に調整すべきである。

本調査は教科の正答率だけでなく、児童生徒質問紙・学校質問紙等のデータも含んでいる。きわめて複雑な多次元データであり、すべての変数間の相関関係等を保持することは難しい。そのため、変数間のクロス分析を行っても、解釈可能な結果が出ないことがあることに留意が必要である。

3.3 データの生成

データ生成については、フリーの統計ソフトウェア HAD⁷を用いて行った。具体的には、以下の手順に従った。これにより、設問間の相関関係を保持した状態で、教科の正答率が真の値±5ポイント程度に収まる疑似データを作成することができた。

- 1) 公立・当日実施データのみ抽出

⁷ HAD は関西学院大学准教授の清水裕士氏が開発した⁷、Excel で動作するフリーの統計ソフトウェアである。データ生成については以下を参考にした。
HAD の便利機能 2 <http://norimune.net/2030>

- 2) 個票データ内の全教科の「正誤区分」(正答・誤答・無解答)を2値データ(1:正答, 0:誤答)に変更, ダミー変数化
- 3) 2)を基に各設問の共分散行列を作成, 整理
- 4) HADにて, 3)で作成した共分散行列を参照して2,000件のデータを生成し, 生成されたデータを2値データ(1:正答, 0:誤答)に調整
- 5) 平均正答数/正答率, 設問別正答率・領域・観点・形式別正答数/正答率を算出
- 6) 1:正答と0:誤答について, 母集団の類型分布を基に, 詳細な正答類型・誤答類型に按分
- 7) 正答数・正答率・学力層及び領域・観点・形式別正答数・正答率を算出

4 今後の展望

4.1 児童生徒データと学校データの紐付け

今回の調査研究では、検討の結果、児童生徒データと学校データを紐付けることはできないものとした。この理由としては、ダミーの学校コードを振って児童生徒データと学校データを紐付けることにより、児童生徒 X と児童生徒 Y が同一学校に所属することがわかってしまうなど、匿名性の度合いが下がってしまうこと、ダミーの学校コードごとに集計した学校の正答率と、学校匿名データに保持されている（学校内の全児童生徒で算出した）平均正答率にずれが生じてしまうこと（これはとくに小規模校において顕著である）、児童生徒データの後ろに学校質問紙回答データがついており、学校質問紙の回答と児童生徒の学力・学習状況を組み合わせた分析は可能であること、学校情報を削除することにより、匿名性を高く保持したままサンプリング率を高める＝データ件数を増やすことができること、等が挙げられる。

一方で、階層線形モデル等の高度なデータ分析を行いたい場合には、学校という集団の特性を考慮することが求められる。以下のような手順で、学校と児童生徒を紐付けた形で匿名データを作成することも考えられるだろう。

- 1) 学校規模＝児童生徒数に合わせて層化し、層ごとに一定割合の学校を抽出する大規模校はトップコーディングする
- 2) 抽出した学校から一部の児童生徒を抽出する（層化二段抽出）
- 3) ダミーの学校コードを振り、同一学校の児童生徒を判別できるようにする

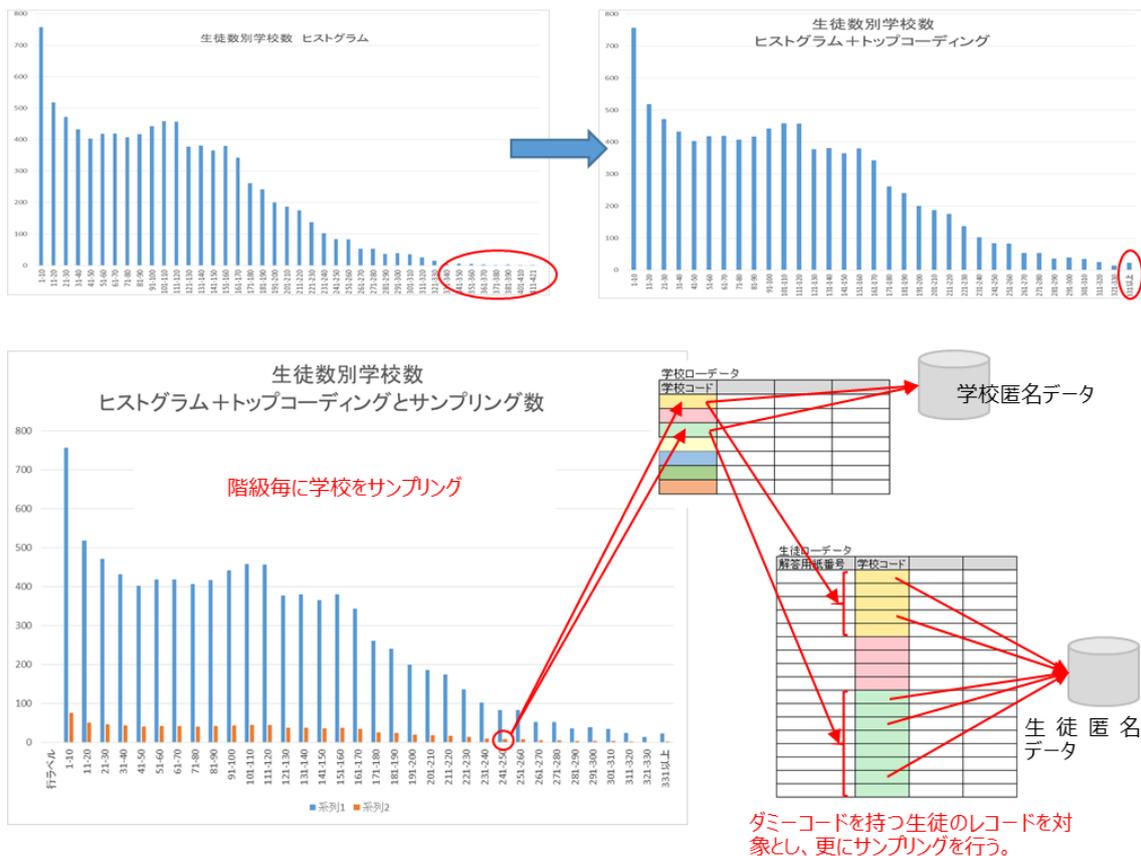


図 5 層化二段抽出のイメージ

4.2 児童生徒データ、学校データの正規化とデータベース化

上記とも関連するが、例えば児童生徒ローデータにおける学校質問紙回答は、学校に所属する児童生徒にすべて同じ内容が入っている。あるいは、学校ローデータにおける児童生徒質問紙の肯定的回答割合のフィールドは、児童生徒質問紙の選択肢別回答割合から算出可能であり、冗長であると思われる。これは、個票データが児童生徒ごとに一人一行、学校ごとに一校一行の行列形式で整備されていることによっている。リレーショナル・データベースの考え方に基づいてデータを正規化し、マスターテーブルとデータテーブル（生徒及び学校）を分割することで、データを引き出しやすい形式にすることも考えられる。

関連して、現在 csv ファイルのヘッダー部分については日本語を含む列名（文字コードは Shift JIS）となっているが、英数字のみのデータとしておいたほうが、多様な分析環境で作業を行うことが可能になると考えられる。テーブル・フィールド・レコード名は基本的に英数字として整備するのがよいと思われる。

これにより、csv形式のファイルとしてデータを提供するだけでなく、データベースをクラウド上に保持し、研究者が分析指示をオンラインで出して、結果の数表だけを得る等、データそのものを提供せずに分析利用を図る等の可能性も開けると思われる。

4.3 パブリックユースデータのバリエーション

疑似データとしてパブリックユースデータを作成する際の課題はデータの真正性である。平均正答率等の一部統計量を保持するとはいえ、児童生徒質問紙・学校質問紙まで含めると、本調査はきわめて複雑な多次元データを含んでおり、疑似データはそのすべての項目間の多変量関係を保持するわけではない。もともと研究利用を想定していないデータではあるが、大学授業等で学習に利用する場合、データの実用可能性に限界があるといえる。

全国消費実態調査の疑似マイクロデータ⁸として公表されているフィールドは非常に少ないこと、また諸外国におけるパブリックユースデータを参照すると、疑似データとして統計表をもとに生成されたものだけでなく、匿名化処理の度合いをさらに高め、データ件数をさらに絞る等、個票データをさらに高度に匿名化したデータをパブリックユースデータとして公開することも考えられる。例えば、平均正答率と読書に関する質問項目のみを抜粋した「読書に関する匿名データ」等、テーマごとに分析に資する形でのパブリックユースデータの生成・公開も考えうるだろう。

⁸ 一般用マイクロデータの利用 <http://www.nstac.go.jp/services/ippan-microdata.html>

資料

有識者委員会開催記録

有識者委員名簿（五十音順・敬称略）

氏名	所属・役職
伊藤 伸介	中央大学 経済学部 教授
後藤 康志	新潟大学 教育・学生支援機構 准教授
福田 幸男	横浜薬科大学 薬学部 教授
南 和宏	統計数理研究所 モデリング研究系 准教授
村山 功 ※主査	静岡大学大学院 教育学研究科 教授

第1回有識者委員会

2017年7月11日 於：株式会社内田洋行

出席委員：伊藤委員・後藤委員・南委員・福田委員・村山委員

議題：「ガイドライン」の共有と意見交換

第2回有識者委員会

2017年9月1日 於：株式会社内田洋行

出席委員：伊藤委員・後藤委員・南委員・福田委員・村山委員

議題：パブリックユースデータと匿名データのあり方（案）について

※匿名データの検討から先に行うことを決定

第3回有識者委員会

2017年10月3日 於：株式会社内田洋行

出席委員：伊藤委員・南委員・福田委員・村山委員

議題：匿名化サンプルデータ①（児童生徒データ）について

技術検討ワーキンググループ

2017年11月29日 於：株式会社内田洋行

出席委員：伊藤委員・南委員

議題：匿名化サンプルデータ②（児童生徒データ・学校データ）について

研究利用ワーキンググループ

2017年12月1日 於：株式会社内田洋行

出席委員：後藤委員・福田委員・村山委員

議題：匿名化サンプルデータ②（児童生徒データ・学校データ）について

第4回有識者委員会

2017年12月12日 於：株式会社内田洋行

出席委員：伊藤委員・後藤委員・南委員・福田委員・村山委員

議題：匿名データの作成手順について

第5回有識者委員会

2018年2月9日 於：株式会社内田洋行

出席委員：伊藤委員・後藤委員・南委員・福田委員・村山委員

議題：匿名データ・パブリックユースデータの作成手順について

匿名データレイアウト

児童生徒匿名データ

項目	データ内容
実施年	※西暦にて表記，数字4桁
地域規模	1: 大都市 2: 中核市 3: その他の市 4: 町村
解答用紙番号	※個票データの解答用紙番号とは異なる番号を連番で付与する
性別	1: 男子 2: 女子 0: 不明
児童生徒質問紙種別	1: 児童生徒質問紙Ⅰ 2: 児童生徒質問紙Ⅱ 3: 児童生徒質問紙Ⅲ ※平成25年度のみ，児童生徒質問紙が分冊で行われたため，当該学校がどの児童生徒質問紙を実施したかを示すもの。以降年度は空欄固定
正答数_ (教科)	
正答率_ (教科)	
学力層_ (教科)	1: A層 2: B層 3: C層 4: D層 ※全国の正答数ヒストグラムを4分割し，正答数が高いほうからA層～D層としたときに当該児童生徒がどの層に当てはまるかを示す

項目	データ内容
	もの。
正答数_ (教科) _ (領域・観点・形式)	
類型_ (設問)	1: 類型 1 2: 類型 2 3: 類型 3 4: 類型 4 5: 類型 5 6: 類型 6 7: 類型 7 8: 類型 8 9: 類型 9 (上記以外の解答) 0: 類型 0 (無解答)
正誤_ (設問)	1: 正答 2: 誤答 0: 無解答
正答率_ (教科) _ (領域・観点・形式)	
児童生徒質問紙回答	1: 選択肢 1 2: 選択肢 2 3: 選択肢 3 4: 選択肢 4 5: 選択肢 5 6: 選択肢 6 7: 選択肢 7 8: 選択肢 8 9: その他 (選択肢以外の回答や複数回答) 0: 無回答
学校質問紙回答	1: 選択肢 1 2: 選択肢 2

項目	データ内容
	3: 選択肢 3 4: 選択肢 4 5: 選択肢 5 6: 選択肢 6 7: 選択肢 7 8: 選択肢 8 9: 選択肢 9 0: その他(選択肢以外の回答や複数回答)・無回答 ※児童生徒が所属する学校の学校質問紙回答を児童生徒 1 行に紐づいて示すもの。

学校匿名データ

項目	データ内容
実施年	※西暦にて表記, 数字 4 桁
地域規模	1: 大都市 2: 中核市 3: その他の市 4: 町村
児童生徒質問紙種別	1: 児童生徒質問紙Ⅰ 2: 児童生徒質問紙Ⅱ 3: 児童生徒質問紙Ⅲ ※平成 25 年度のみ, 児童生徒質問紙が分冊で行われたため, 当該学校がどの児童生徒質問紙を実施したかを示すもの。以降年度は空欄固定
正答数_ (教科)	※平均正答数
正答率_ (教科)	※平均正答率 (整数値に四捨五入)

項目	データ内容
第1四分位_(教科) 第2四分位_(教科) 第3四分位_(教科) 第4四分位_(教科)	※教科の正答数を基準に、学校内をA層～D層の4層に分割したときに区分点となる正答数の値。
最頻値_(教科)	※学校内の正答数の最頻値を示すもの。
標準偏差_(教科)	※平均正答数の標準偏差
正答数_(教科)_(領域・観点・形式)	※領域・観点・形式別平均正答数
正答率_(教科)_(領域・観点・形式)	※領域・観点・形式別平均正答率(整数値に四捨五入)
無解答数_(教科)	※平均無解答数
無解答率_(教科)	※平均無解答率
無解答数_(教科)_記述式	※記述式のみ無解答数を示すもの。
無解答率_(教科)_記述式	※記述式のみ無解答率を示すもの。
割合_(教科)_A層 割合_(教科)_B層 割合_(教科)_C層 割合_(教科)_D層	※全国の正答数ヒストグラムをもとに算出したA層～D層(学校のA層～D層とは異なる)に、当該学校の児童生徒が当てはまる割合を示すもの。
児童生徒質問紙回答割合_(質問番号)	※児童生徒質問紙の当該学校の肯定的回答割合を示すもの。
児童生徒質問紙回答割合_当日実施のみ_(質問番号)_(選択肢番号)	※児童生徒質問紙の当該学校の質問ごと・選択肢ごとの回答割合を示すもの。
学校質問紙回答_(質問番号)	
結果チャート_全国_学校質問紙_(領域) 結果チャート_全国_児童生徒質問紙_(領域)	※全国基準の「結果チャート」の得点を示すもの。
結果チャート_県_学校質問紙_(領域) 結果チャート_県_児童生徒質問紙_(領域)	※都道府県基準の「結果チャート」の特典を示すもの。

調査研究体制

調査研究責任者：

- ・伊藤博康（株式会社内田洋行 教育総合研究所 研究開発部長）

研究担当：

- ・平野智紀（同 教育総合研究所 主任研究員）
- ・加藤紗夕理（同 教育総合研究所）

技術担当：

- ・杉山知之（同 AP&プラットフォーム開発部 ラーニングプラットフォーム開発課長）
- ・岡元英公（同 AP&プラットフォーム開発部
ラーニングプラットフォーム開発課 専門課長）
- ・相沢優（同 AP&プラットフォーム開発部 ラーニングプラットフォーム開発課）

事務局：

- ・宍倉佳苗（同 教育総合研究所）