

令和7年度総務省発注請負調査

「全国学力・学習状況調査結果の長期的分析等に係る調査研究の請負」

# データクレンジング及び経年データ作成に関する 手引き

三菱 UFJ リサーチ&コンサルティング株式会社

## もくじ

|   |    |
|---|----|
| もくじ   | 2  |
| 1. はじめに                                     | 3  |
| 1-1. 本手引きの内容と対象                             | 3  |
| 1-2. データクレンジングの必要性とメリット                     | 5  |
| 1-3. 経年データ作成のメリット                           | 5  |
| 1-4. データの取扱いにおける重要な3つの視点                    | 7  |
| 1-5. 本手引きで扱うデータクレンジングの全体像                   | 9  |
| 1-5-1. ①基本的なデータクレンジング                       | 9  |
| 1-5-2. ②経年データ作成のためのクレンジング                   | 9  |
| 1-6. 本手引きで想定するツール                           | 10 |
| 1-7. 手引きの構成                                 | 12 |
| 2. 基本的なデータクレンジング                            | 15 |
| 2-1. 一時点データの構造を把握する                         | 15 |
| 2-2. 基本的なデータクレンジングを行う                       | 17 |
| 2-3. 経年データの活用目的を検討する                        | 20 |
| 2-4. 経年データの作成に適したデータであるか確認する                | 21 |
| 3. 調査項目をつなげるためのデータクレンジング                    | 25 |
| 3-1. 「調査項目マスタ」を作成する                         | 25 |
| 3-1-1. 手順①：複数時点の調査項目情報の一覧表を Excel で作成する     | 25 |
| 3-1-2. 手順②：複数時点の調査項目情報の一覧表から「調査項目マスタ」を作成する  | 30 |
| 3-2. 調査項目ごとに識別 ID を付与する                     | 42 |
| 3-2-1. 手順③：調査項目の識別 ID の採番ルールを決定し、各調査項目に付与する | 42 |
| 3-2-2. 手順④：調査項目マスタを用いて変数対応表を作成する            | 45 |
| 3-2-3. 手順⑤：変数対応表を用いてデータセットに調査項目の識別 ID を反映する | 46 |
| 4. 調査対象をつなげるためのデータクレンジング                    | 47 |
| 4-1. 調査対象の接続に必要な各種接続キーを整理、統一する              | 47 |
| 4-1-1. 手順⑥：一貫した「接続キー」を設定し、データセットに反映する       | 49 |
| 4-1-2. 手順⑦：接続単位ごとのデータを作成する                  | 50 |
| 4-2. 経年データを作成する                             | 56 |
| 4-2-1. 手順⑧：接続キーを用いてデータを接続する                 | 56 |
| 4-2-2. 手順⑨：接続したデータのチェック（接続キー）を行う            | 58 |
| 4-2-3. 手順⑩：接続したデータのチェック（データエラー）を行う          | 59 |
| 5. 経年データを用いて分析を行う                           | 61 |
| データクレンジング及び経年データ作成のチェックリスト                  | 63 |
| ■ 参考資料                                      | 65 |

# 1. はじめに

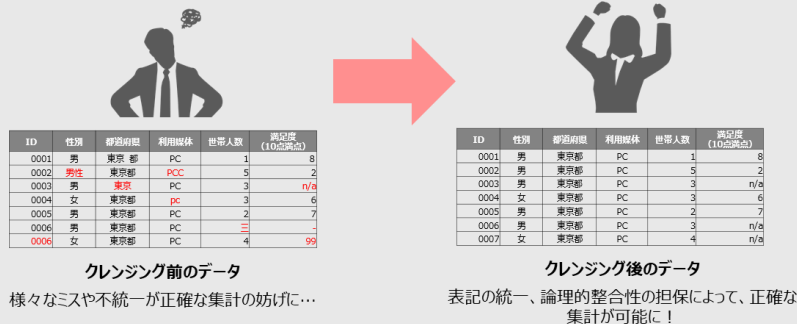
## 1-1. 本手引きの内容と対象

- 本手引きは、個人や企業、各種団体、自治体等を対象としたアンケート調査などの各種個票データ（調査対象者ごとの回答結果が一覧で入力されたデータ）を分析・活用する際に必要となる、「データクレンジング」の基本的考え方や留意点をまとめたものです。
- また、本手引きでは、毎年又は数年おきなど、定期的に同じ対象（母集団）に対して繰り返し実施される調査における複数の時点のデータを接続した「経年データ」を作成する方法も説明しています。

### ✔ Point 本手引きの主な対象者例

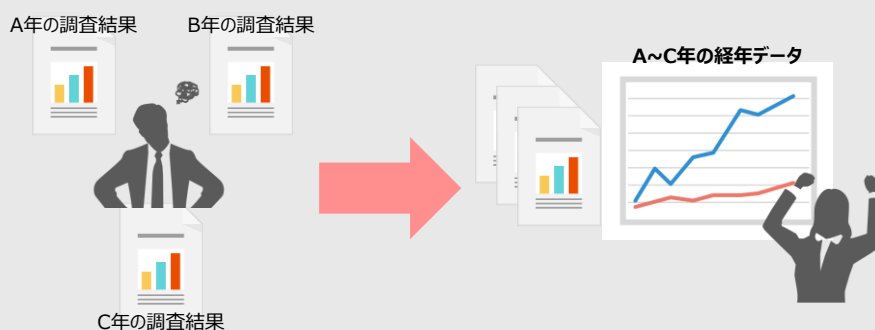
- 所属組織等が実施する調査（アンケート調査、行政記録を用いた調査等）の実施担当になっ方（初めてデータを取り扱う方にも分かりやすい記載を心がけています。）
- 所属組織等が保有するアンケート調査等を分析・活用したいと考えている方

＜取得した個票データをクレンジングし、正確な集計による実態把握に活用しましょう！＞



- 所属組織等が実施する定期的な調査等で得られるデータを複数時点分接続して、変化や傾向などの実態を把握したい方、更に課題の発見などを行いたいと考えている方

＜経年データを作成し、変化や傾向の把握、課題発見等に活用しましょう！＞





### ✔ Point 個票データの活用、こんな困りごとはありませんか？

- アンケートの個票データを確認したところ、同じ回答でも表記が異なっていたり、入力方法が異なっていたりするなど、そのまま集計してよいか不安がある。
- 手始めにアンケートを集計してみたが、正しい集計値が出ているのか不安がある。
- 手元に複数年分のアンケートの個票データがあるが、各時点の調査で少しずつ質問文や選択肢が変わっているところがあり、その変更をどのように扱えばよいか分からない。

本手引きでは、データクレンジング及び経年データの作成方法の説明を通じて、上記のような困りごと解消の一助となることを目的としています。

### ✔ Point 本手引きの前提条件について

本手引きは、令和7年度総務省発注請負調査「全国学力・学習状況調査結果の長期的分析等に係る調査研究の請負」における、総務省・文部科学省による実証的共同研究の成果をもとに取りまとめられたものです（以下、本調査研究に触れる際は「実証的共同研究」と表記）。具体的には、文部科学省が毎年小学校第6学年・中学校第3学年を対象に実施している「全国学力・学習状況調査」の個票データの貸与を受け、データのクレンジング、経年データの作成、分析を行った工程をもとに、それを一般化、抽象化したり、一部具体例として用いる形で、データクレンジングや経年データ作成手順を取りまとめています。

なお、上記実証的共同研究報告書では、全国学力・学習状況調査の個票データに対する具体的なデータクレンジング等の内容を詳述しています。

#### 【参考：「全国学力・学習状況調査」について】

文部科学省では、平成19年度から小学校第6学年、中学校第3学年の原則全児童生徒を対象として、全国的な児童生徒の学力や学習状況の把握による教育施策の検証・改善や学習指導の充実・改善に役立てることを目的として「全国学力・学習状況調査」を実施しています。

全国学力・学習状況調査の結果は大規模に長年実施されている貴重なデータですが、複数年度にわたる調査項目の整理やデータクレンジングの必要性、個人の追跡ができない設計となっている点、また令和6年度以前の悉皆調査ではIRT（項目反応理論）による教科調査の比較可能性が担保されていない点などの課題があり、時系列的なデータの接続が難しいという課題がありました。（なお、令和7年度からは全国学力・学習状況調査の悉皆調査にIRT（項目応答理論）が導入されています。）

上記実証的共同研究では平成19年度～令和6年度における全国学力・学習状況調査の解答（回答）データを用い、上記課題に対処しながら調査データの接続を行ったプロセス及び分析結果を取りまとめています。

## 1-2. データクレンジングの必要性とメリット

- アンケート調査等によって取得した個票データは、すぐに活用できるとは限りません。
- データの活用に当たっては、データの誤りや不整合、重複、欠損などを修正・除去し、分析に適した状態に整える一連の作業である「**基本的なデータクレンジング**」の過程を経て、はじめて正確に活用できるようになります。
- データの活用に先立ち、適切に基本的なデータクレンジングを行うことで、分析結果の精度向上、誤った解釈の防止など、実務上大きな効果が得られます。
  - ➡詳細は「2-2.基本的なデータクレンジングを行う」を参照
- また、定期的に同じ対象（母集団）に対して繰り返し実施される調査のデータを複数時点で接続し、経年データを作成する際にも、データクレンジングを行う必要があります。なぜなら、たとえ**定期的**に実施される調査であっても、その調査内容が毎回同一であるとは限らないからです。実際には、設問文や選択肢などが変更されていることがよくあります。経年データを作成する際には、まず「活用したい個票データが経年データとするのに適しているか」、また、「データの利用目的（経年データ分析の目的や、変化を確認したい点等）に照らして、経年データとする意義のあるデータか」を確認する必要があります。その上で、経年データを作成するためのデータクレンジングを行う必要があります。これらの工程により、**正確性が確保された経年データを活用できるよう**になります。
  - ➡詳細は「2-4. 経年データの作成に適したデータであるか確認する」を参照

## 1-3. 経年データ作成のメリット

- 複数時点における調査データを接続した経年データを作成することで、「**数値の変化や関係を把握**」でき、そして「**時間的な変化を踏まえた分析結果の解釈**」が可能になります。さらに、様々なデータの可視化に強みを持つ Power BI などのツールと組み合わせると、分析結果の理解や共有がより容易になります。
- また、作成した経年データを、因果推論のための統計的手法を用いて分析することで、データ間の「**因果関係を予測する**」ことが可能となるなどのメリットが得られます。経年データを用いると、一時点で発生したある変数（原因と考えられる要素や政策的介入等）の変化が、その後、他の変数にどのような影響を与えるのかを数値的变化によって把握し、解釈することが可能になります。このことは、政策効果の内容や程度・影響範囲の把握を行い、政策の振り返りや改善の検討に役立ちます<sup>1</sup>。

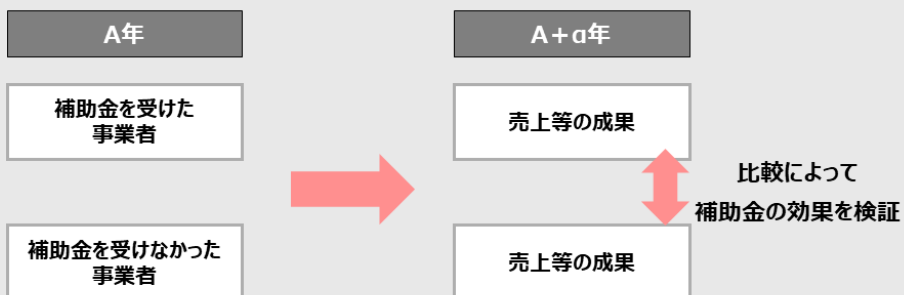
---

<sup>1</sup> ただし、経年データの作成に当たっては、複数時点にわたる調査であることが前提であるため、各時点の調査対象が同じであるか、経年変化の要因の特定に誤りはないか（因果関係ではなく相関関係となっている可能性はないか）といったことなどについて、十分に留意した上で分析に臨むことが必要です。

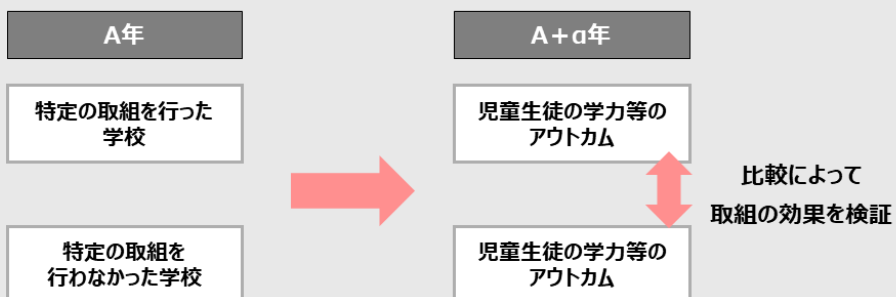
✔ Point 因果関係を予測する

経年データの作成により、ある時点（下の図表では A 年）での施策・介入（原因）と、その後（下の図表では A + α 年）の結果の間の関係性を比較分析・検証する「因果推論」のための基盤となるデータを作成することができます。

例①：事業者に対する補助金の効果分析



例②：学校における特定の取組の効果分析



## 1-4. データの取扱いにおける重要な3つの視点

- 本手引きで紹介するデータクレンジング及び経年データの作成に係る具体的な手順の案内に先立ち、作業全体を通じた**データの取扱いにおける重要な3つの視点**を紹介します。これらの視点は、信頼性・正確性の高いデータを効率的に整備・管理する際の重要なポイントとなります。

### ✔ Point データの取扱いにおける3つの視点

#### ① 一貫性

複数時点のデータを比較・接続できるようにするため、まず異なる時点間で共通するルールを定める。これによって、設問文や選択肢の意味、表現・表記、ID・コード体系や記録・管理方法などに一貫性を確保する。

#### ② 識別可能性

調査対象や分析対象を明確に識別できるよう、固有の情報（ID やコード）を付与する。

#### ③ 再現性

調査担当者や担当部局が変わっても、データクレンジングやデータの紐づけの過程を再現・追跡できるよう、調査票やデータ整備・管理に関する方針・判断基準・ルール・情報を文書化しておく。また、データクレンジング作業を行う際は複製したファイルを用いて作業前のファイルを残すなど、データクレンジングの経緯が遡れるようファイルを適切に保存しておく。

### ✔ Point 調査設計に役立つ参考資料

上述の3つの視点は、取得済みデータのクレンジング時のみならず、**調査設計の段階から意識することによって**、信頼性・正確性の高いデータが収集可能となり、データの活用や経年データの作成が非常にスムーズになります。

次章から説明する具体的なデータクレンジング作業についても、その手順を調査設計段階から認識し、データの収集方法や、データクレンジングの経緯の記録方法を工夫することで、ミスや不整合の発生を予防し、データクレンジングに係るコストを抑えることが可能です。

本手引きでは主に取得済みのデータの整備・管理方法について扱いますが、以下に示す参考資料には、いずれも調査設計段階における検討事項について、一貫性、識別可能性、再現性の観点に基づき説明されています。調査設計や改訂について検討される際には、ぜひご参照ください。

- **総務省統計局「なるほど統計学園」**

(<https://www.stat.go.jp/naruhodo/index.html>)

調査の計画（目的・対象・時期・方法など）から調査票の作成まで、調査設計に必要な検討事項が具体的な例とともに示されています。

- **経済産業省 大臣官房調査統計グループ 統計企画室「令和4年度統計調査業務改善に関する調査研究事業 調査報告書（1）要約版資料の作成」**

([https://www.meti.go.jp/meti\\_lib/report/2022FY/000670.pdf](https://www.meti.go.jp/meti_lib/report/2022FY/000670.pdf))

統計作成プロセスを企画、実査、審査・集計、公表・提供、評価、マネジメント、民間委託、地方委任の8つに分類し、政府統計職員が統計作成における業務マニュアル・手順書等の整備時に必要な項目がまとめられています。

- **総務省統計局「品質評価事項チェックリスト」**

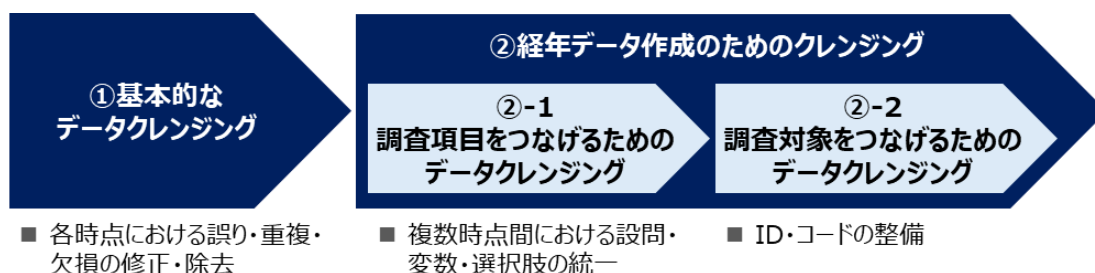
(<https://www.stat.go.jp/data/guide/pdf/checklist26.pdf>)

調査設計・実施・公表に際して検討すべき内容が、チェックリストとして整理されています。

## 1-5. 本手引きで扱うデータクレンジングの全体像

- 前述の通り、本手引きで扱う「データクレンジング」とは、
  - ① 基本的なデータクレンジング（誤り、不整合、重複、欠損などを修正・除去）
  - ② 経年データ作成のためのクレンジング（調査項目をつなげるための設問文等の統一や調査対象をつなげるための識別情報の付与・統一）

の双方を含むものであり、経年分析や活用に適した状態に整えるための作業を指しています。



### 1-5-1. ①基本的なデータクレンジング

- 経年データの作成有無にかかわらず、まずは一時点の調査データにおいて、データの誤り、不整合、重複、欠損などを修正・除去し、**正確性のある状態にする作業**が必要となります。

### 1-5-2. ②経年データ作成のためのクレンジング

- 複数時点の調査データの接続を行う場合、①の基本的なデータクレンジングを行っても、時点間で設問文や選択肢の表記、IDの付与方法など様々な差異が生じている可能性があります。また、分析やデータ接続の目的に応じた調査項目や識別情報の整理が必要になります。これらを放置すると、**経年データの作成や経年分析に多大な手間や誤りが発生するリスク**があります。
- 経年データ作成のためのクレンジングについて、本手引きでは以下の2種類に分けた上で手順を示しています。

#### ②-1 調査項目をつなげるためのデータクレンジング

- 複数時点で行われた調査では、時点間で設問文や選択肢の表記が統一されていなかったり、同じ趣旨の調査項目と思われるものであっても、その「内容の差異」が生じている場合があります。正しく接続できるよう、同じ調査項目の設問文や選択肢の表記を統一したり、「内容の差異」がある調査項目を同一と見なすか、あるいは別の調査項目とみなすか等の対応方針を決定していく必要があります。

## ②-2 調査対象をつなげるためのデータクレンジング

- 複数時点で取得した調査データの接続に当たっては、接続の単位（例：個人、学校、市町村など）ごとに、同じ対象に同じ識別情報（回答者に与えられる ID やコード）を一貫して与えること、つまり「接続キー」を揃える必要があります。接続キーによって異なる取得時点でも確実に同一の調査対象を特定でき、データを正確に接続することが可能になります。

### <経年データ作成のためのクレンジング前後のデータのイメージ>

#### <クレンジング前データ>

| 年度    | 学校名  | 学級数(1: はい/2: いいえ)      | 回答 | 席替え (実数回答[回])              | 回答 |
|-------|------|------------------------|----|----------------------------|----|
| 2019年 | A小   | (2) 学級数はこの1年で増加していますか  | 1  | (3) 1年間で、席替えを何回実施していますか    | 2  |
| 2020年 | A小学校 | (2) 学級の数はこの1年で増加していますか | 2  | (3) 1年間で、席替えをどのくらい実施していますか | 3  |
| 2021年 | A小学校 | (1) 学級数はこの1年で増加していますか  | 2  | (2) 1年間で、席替えを何回実施していますか    | 2  |

#### ②-2 調査対象をつなげる

= 接続・分析単位での識別情報の付与・統一

- 学校名：「A小」「A小学校」→「学校ID:001」と「A小学校」に統一

#### ②-1 調査項目をつなげる

= 設問や変数名、選択肢などの統一

- 「学級数/学級の数」→「学級数」、「何回/どのくらい」→「何回」に統一
- 学級数・席替えの設問について、それぞれ設問番号を統一

#### <クレンジング後データ>

項目・対象が揃うと、年度を跨いだ比較・分析が可能になります。

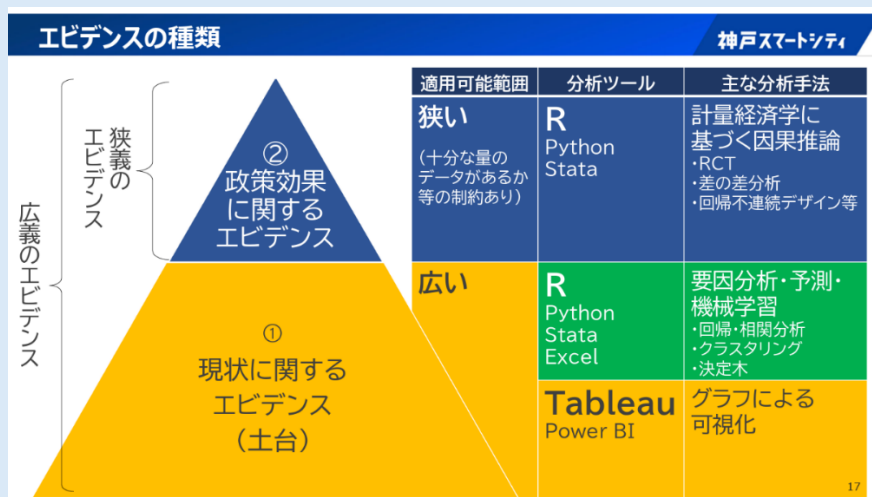
| 年度    | 学校ID | 学校名  | (2) 学級数はこの1年で増加していますか | (3) 1年間で、席替えを何回実施していますか |
|-------|------|------|-----------------------|-------------------------|
| 2019年 | 001  | A小学校 | 1                     | 2                       |
| 2020年 | 001  | A小学校 | 2                     | 3                       |
| 2021年 | 001  | A小学校 | 2                     | 2                       |

## 1-6. 本手引きで想定するツール

- 本手引きでは以降、具体的なデータクレンジングの手順について扱います。基本的には、多く組織で使用されているであろう **Microsoft Excel 等の表計算ソフト** に記録された個票データを想定した説明を行っています。
- 一方、Excel での動作が重くなるほどデータ量が多い（数万行以上）場合は、R や Python、SPSS などの統計ソフトを用いることで、より正確かつ効率的なデータクレンジング及び経年データの作成、分析作業が可能となります。
- その他、各種分析ツールの特徴については、以下をご参照ください。

Point データ利活用の参考ツール

- 国連統計委員会 世界統計の日フォーラム 2025 講演資料  
神戸市「神戸市における EBPM の実践 ～日本の地方行政におけるビッグデータ活用～」  
p.17  
([https://www.soumu.go.jp/main\\_content/001043825.pdf](https://www.soumu.go.jp/main_content/001043825.pdf))

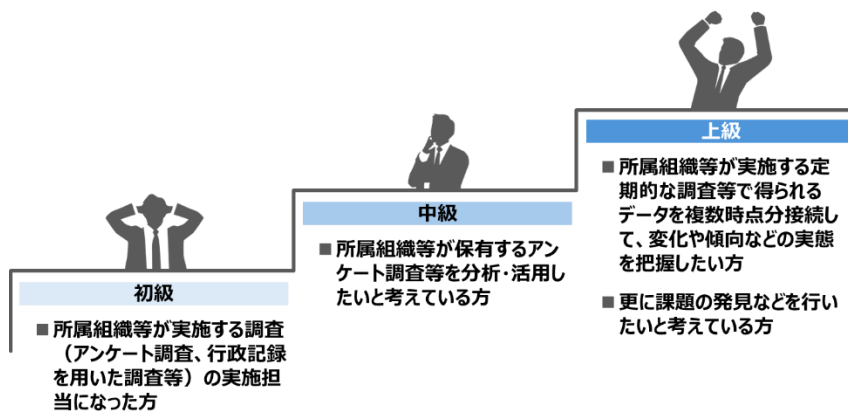


- デジタル庁 オープンデータ取組の質評価指標（ベータ版）・研修資料  
「オープンデータ研修テキスト 中級編（別添資料）データ利活用の参考ツール」  
([https://www.digital.go.jp/assets/contents/node/basic\\_page/field\\_ref\\_resources/651e3f74-9a8e-4797-98e9-97f17b0dcce7/343cf1bf/20250724\\_resources\\_open\\_data\\_materials-for-le.pdf](https://www.digital.go.jp/assets/contents/node/basic_page/field_ref_resources/651e3f74-9a8e-4797-98e9-97f17b0dcce7/343cf1bf/20250724_resources_open_data_materials-for-le.pdf))

## 1-7. 手引きの構成

- 本手引きは 5 章構成となっています。
- 第 2 章では基本的なデータクレンジングの手順やポイントを説明します。章の後半では、経年データの作成に向けたチェックリストを設けています。
- 続いて、経年データ作成のためのクレンジングとして、第 3 章では調査項目をつなげるためのデータクレンジング、第 4 章では調査対象をつなげるためのデータクレンジングについて、それぞれの章で具体的な手順やポイントを説明します。また、第 4 章の後半では、データクレンジングが完了したデータの接続方法について説明します。
- 第 5 章では、経年データをどのように活用できるかについて「全国学力・学習状況調査」の実際の分析事例を用いて紹介します。経年データの活用イメージを持ちたい方は、先に 5 章をお読みください。
- 本手引きを活用して、お持ちのデータを、信頼性が高く、データ分析の観点からも取り扱いやすいものにしていただきたいと思います。そして、データを利用した問題解決が各組織で一層推進されることを期待しています。
- 以下の図や次頁のフローチャートを参考にしながら、データ利活用に関する習熟度や必要に応じて、該当する章をご確認ください。

### <データ利活用に関する習熟度>



### <習熟度に応じて優先的に参照されたい項目>

| 章  | 2. 基本的なデータクレンジング |     |     |     | 3. 調査項目をつなげるためのデータクレンジング |       |     | 4. 調査対象をつなげるためのデータクレンジング |     | 5. 分析 |
|----|------------------|-----|-----|-----|--------------------------|-------|-----|--------------------------|-----|-------|
|    | 2-1              | 2-2 | 2-3 | 2-4 | 3-1-1                    | 3-1-2 | 3-2 | 4-1                      | 4-2 |       |
| 初級 | ○                | ○   | ○   | ○   | ○                        | ○     | ○   | ○                        | ○   | ○     |
| 中級 | 必要に応じて参照         |     | ○   | ○   | ○                        | ○     | ○   | ○                        | ○   | ○     |
| 上級 | 必要に応じて参照         |     |     |     | ○                        | ○     | ○   | ○                        | ○   | ○     |

## ＜データクレンジング及び経年データ作成のためのフローチャート＞

### 2. 基本的なデータクレンジング

| 章/節/項 | ニーズ                                 | 概要   |
|-------|-------------------------------------|--|
| 2-1   | 一時点のデータ構造を把握したい<br>→p.15            | <ul style="list-style-type: none"> <li>✓ 多くの調査データには、調査項目に対する回答のほか、調査対象を識別する ID や属性情報などが含まれるといった共通した構造が見られます。</li> <li>✓ この節では、このようなデータ構造の特徴を確認し、その後のデータクレンジングを効率的かつ正確に行えるように備えます。</li> </ul>   |
| 2-2   | 一時点のデータを対象としたクレンジングを行いたい<br>→p.17   | <ul style="list-style-type: none"> <li>✓ 単年度分など、一時点の調査データにおいて、そのデータ内にデータの誤り、不整合、重複、欠損などが見られた場合には、修正・除去し、時点間で統一された状態にします。</li> <li>✓ この節では、修正・除去の具体的な方法を説明します。</li> </ul>   |
| 2-3   | 「経年データ」の活用目的や作業方針を検討したい<br>→p.20    | <ul style="list-style-type: none"> <li>✓ 一時点データの基本的なデータクレンジングを行った後、さらに経年データの作成に進む場合は、まずは経年データの活用目的を明確にする必要があります。</li> <li>✓ この節では、その目的に沿ってデータクレンジング方針を整理し、効率的に作業を進める方法を説明します。</li> </ul>   |
| 2-4   | データが「経年データ」の作成に適しているか確認したい<br>→p.21 | <ul style="list-style-type: none"> <li>✓ 各種調査等によって取得したデータには、様々な状態のものが存在します。集計対象や、設問文・選択肢の変更、識別情報の有無、欠損の有無を確認することで、経年データの作成可能性を判断することができます。</li> <li>✓ この節では、経年データの作成に関する基礎知識を理解した上で、お持ちの調査データが経年データの作成に適しているか確認するためのポイントを説明します。</li> </ul> |

### 3. 調査項目をつなげるためのデータクレンジング

|       |                                 |  |
|-------|---------------------------------|--|
| 3-1-1 | 調査項目（設問文・選択肢）を一覧化したい<br>→p.25   | <ul style="list-style-type: none"> <li>✓ 複数時点のデータをつなげるには、接続対象とする期間の中で調査項目の意味や構造が一致しているかを確認する必要があります。しかし、実際には設問文や選択肢が調査時点ごとに異なることが多いため、まず「調査項目マスタ」を作成し、調査項目の対応関係を確認できるようにします。</li> <li>✓ この項では、「調査項目マスタ」の作成方法を説明します。</li> </ul> |
| 3-1-2 | 調査項目（設問文・選択肢）の変化を整理したい<br>→p.30 | <ul style="list-style-type: none"> <li>✓ 「調査項目マスタ」によって一覧化すると、時点間での設問文や選択肢の違いが分かりやすくなります。経年データを作成するにあたっては、最初に一時点データを整理したのと同様に、この違いを再度整理する必要があります。</li> <li>✓ この項では、違いを整理する際の具体的な手順を説明します。</li> </ul>                            |
| 3-2   | 調査項目ごとに識別 ID を付与する方法を知りたい       | <ul style="list-style-type: none"> <li>✓ 各時点でのデータが整ったら、それぞれを繋ぎ合わせて、一つの大きなデータにする必要があります。そのための鍵を握るのが「識別 ID」で</li> </ul>  |

⇒p.42

す。データの参照・集計を円滑に行えるよう、各調査項目に固有の識別 ID を付与する必要があります。

- ✓ この節では、調査項目の識別 ID の採番ルールを決め、各調査項目に識別 ID を付与する方法を説明します。

#### 4. 調査対象をつなげるためのデータクレンジング

4-1 複数時点で統一された「接続キー」を整理したい

⇒p.47

- ✓ データを接続する際には、回答データを紐づける接続単位ごとに、「接続キー」が必要になります。複数時点で統一した接続キーを用意しましょう。

- ✓ この節では、接続キーの考え方や設定方法を説明します。

4-2 各時点のデータを経年接続したい

⇒p.56

- ✓ 接続キーを用いて、調査対象（接続単位）をつなげます。正しく接続できたかを確認するために、データの数え上げや記述統計量からエラーがあるかをチェックし、エラーがあれば修正を行います。

- ✓ この節では、どのようにデータを接続し、エラーを修正するかを説明します。

#### 5. 経年データを用いた分析

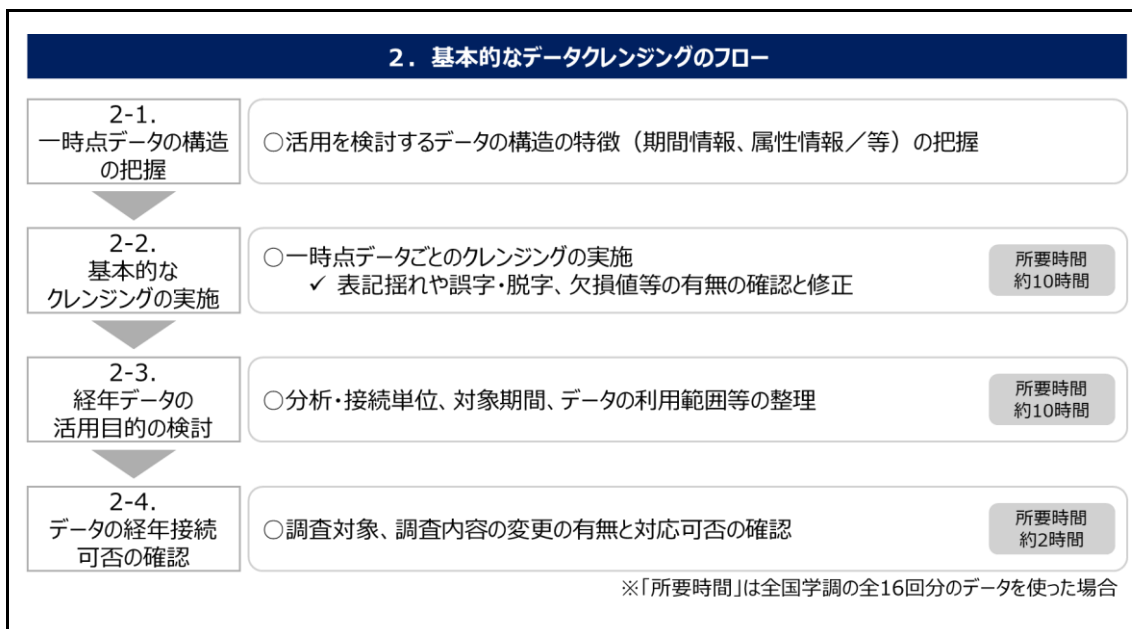
5 経年データを用いた分析の事例を知りたい

⇒p.61

- ✓ この章では、経年データを用いた分析手法を紹介するとともに、「全国学力・学習状況調査」の経年データを用いた分析事例を紹介します。

## 2. 基本的なデータクレンジング

この章ではまず、任意の一時点の調査データについて、そのデータ構造を把握した上で、基本的なデータクレンジングの手順やポイントについて扱います。その上で章の後半では、経年データを作成する場合に必要な事前準備のプロセスについて紹介します。



### 2-1. 一時点データの構造を把握する

|    |  |
|----|--|
| 目的 | まずは一時点のデータの共通した構造を把握する   |
|    | <p>✓ Point</p> <p>多くの調査データにおいては、調査項目に対する回答のほか、調査対象を識別するID やコード<sup>2</sup>、属性情報などが含まれ、共通した構造が見られます。まずは、このようなデータ構造の特徴を確認し、その後のデータクレンジングを効率的かつ正確に行えるように備えましょう。</p> |

- 多くの調査データには、主に以下のような情報が含まれています。データを活用する場合は、まずはこのようなデータ構造の特徴を確認しておくようにしましょう。こうした全体の見通しを得ておくことにより、

<sup>2</sup> 経年データを作成する際に必要になるため、データセットに含まれていない場合は、「4-1. 調査対象の接続に必要な各種接続キーを整理、統一する」をご確認ください。

データクレンジングが必要な箇所の特定が進めやすくなるなどのメリットがあります。

|                           |  |
|---------------------------|--|
| ①回答取得<br>(調査) 時点・<br>期間情報 | 調査実施年度や月など、回答取得（調査）時点・期間を示す情報です。<br>次章で扱う経年データ作成時には必須の情報となります。   |
| ②識別情報                     | 調査対象（個人、世帯、施設など）を一意に識別するためのIDや番号です。<br>匿名性を保ちつつ、経年でのデータ接続や追跡にも利用されます。<br>(識別情報が含まれていない場合は第4章参照)        |
| ③属性情報                     | 居住地域（都道府県、市区町村など）や年齢、性別、職業など、調査対象の基本的な特徴です。例えば全国調査では、一般的に地域情報が含まれ、個人を調査対象とした場合は、さらに詳細な属性が収集されることがあります。 |
| ④回答情報                     | 調査の各調査項目に対する回答内容を記録した回答情報です。選択肢番号や自由記述式、数値など、調査項目の内容に応じた形式でデータが蓄積されます。                                 |



### データ構造の例

- 学校を対象とした質問調査のローデータイメージ

| 実施年  | 学校<br>コード | 学校名  | 都道府県<br>コード | 都道府県<br>名 | 国公私 | 学校種 | ... | 設問<br>1 | 設問<br>2 | 設問<br>3 | ... |
|------|-----------|------|-------------|-----------|-----|-----|-----|---------|---------|---------|-----|
| 2025 | 10000     | A小学校 | ××××        | □□県       | 公立  | 小学校 |     | 1       | 2       | 4       |     |
| 2025 | 10001     | B小学校 | ××××        | □□県       | 公立  | 小学校 |     | 3       | 4       | 5       |     |
| 2025 | 10002     | C小学校 | ××××        | □□県       | 私立  | 小学校 |     | 4       | 2       | 1       |     |


**①回答取得時点  
情報**  
データが取られた  
時点を表す

**②識別情報**  
回答者に応じて  
与えられるID

**③属性情報**  
回答者の基本的  
な属性を表す

**④回答情報**  
調査で得られた  
回答データ

## 2-2. 基本的なデータクレンジングを行う

|    |  |
|----|--|
| 目的 | 一時点データを対象として、基本的なデータクレンジングを行う  |
|    |  <b>Point</b><br>任意の一時点の調査データにおいて、そのデータ内にあるデータの誤り、不整合、重複、欠損などを修正・除去し、時点間で統一された状態にします。 |

- 任意の一時点の調査データ内には、各回答者による記載のゆれや、複数のデータ管理者による記録方法のゆれなどといった様々な要因により、同じ意味を持つ回答等であっても記載方法が異なっている場合があります。これにより、**集計時に機械判読する際、本来は同じ意味を持つ回答が別の回答・データとして扱われてしまう**など、正確な集計に支障をもたらします。こうした事態を防ぐために、**集計前に基本的なデータクレンジングを行う必要**があります。
- なお、再現性の確保のため、行ったデータクレンジングの方針については文書化しておくことに加え、データクレンジング前の元ファイルは残して、複製したファイルを用いて作業を行うようにしましょう。
- また、クレンジングを行う複製ファイルにおいては、**以後の集計や分析に不要な個人情報に関するデータは削除しておく**ようにしましょう。
- 基本的なデータクレンジングにおいて、データの誤り・不整合・重複・欠損などを発見、修正するために確認すべき主要なポイントは以下の通りです。

|             |   |
|-------------|---|
| 全角・半角の統一    | 例：「1（全角）」と「1（半角）」<br>→いずれかに統一する必要があります。特に数値の場合は、機械判読性の観点から半角に統一することが望ましいでしょう。       |
| 表記ゆれの統一     | 例：「東京都」と「東京」<br>→表記の正確性や、他の回答との識別可能性などの観点から、いずれかに統一する必要があります。                       |
| 漢数字と算用数字の統一 | 例：「1時間」と「一時間」<br>→数値データとして集計する場合は、単に「1（半角）」に統一します。集計を効率的に行うためにもデータ入力時に数字での入力を推奨します。 |
| 大文字・小文字の統一  | 例：「PC」と「pc」<br>→表記の正確性などの観点から、いずれかに統一する必要があります。                                     |
| 誤字・脱字の修正    | 例：「保健」と「保険」<br>→正確な表記に統一する必要があります。特に自由記述回答の場合で、正確な表記が一意に特定できない時は、原文のままとする場合もあります。   |
| 重複データの削除    | 例：同じ回答者 ID が複数存在する場合<br>→例の場合、2つのケースが考えられます。  |

|                             |   |
|-----------------------------|---|
|                             | <p>①別々の回答者に同じIDが割り振られている場合、こうした事象が生じた原因を調査した上で、どちらかに別のIDを割り当てます。</p> <p>②同じIDの回答者による重複回答が確認できる場合、こうした事象が生じた原因を調査した上で、「時系列上、先に（後に）回答されたものを『正』とする」「無回答箇所の少ない回答を『正』とする」などの方針を定めて、どちらかの回答を残し、もう一方を削除します。</p>  |
| <p>欠損値（無回答や非該当）表記の統一と識別</p> | <p><u>例：無回答箇所について、「99」「-」「n/a」などの表記ゆれがある</u><br/>→使用する分析ソフトのルールや、通常の回答と重複しない方法等を考慮し、いずれかに統一する必要があります。</p> <p><u>例：無回答（回答可能な調査項目に回答していない）と、非該当（回答する必要のない調査項目に回答していない）が識別されていない</u><br/>→お使いの分析ソフトのルールや、通常の回答と重複しない方法等を考慮した上で、無回答と非該当は別々の表記ルールで区別する必要があります。</p> |

- 上記に関する詳しい内容や具体的なデータクレンジング方法については、以下をご参照ください。

|   |
|---|
| <ul style="list-style-type: none"> <li>● <b>Excel を用いる場合</b> <ul style="list-style-type: none"> <li>➢ 千葉大学アカデミック・リンク・センター「データクリーニング入門（初心者向け）」<br/>(<a href="https://alc.chiba-u.jp/eyr/2025/09/24/data-cleaning.html">https://alc.chiba-u.jp/eyr/2025/09/24/data-cleaning.html</a>)</li> <li>➢ 東京大学数理・情報教育研究センター「数理・データサイエンス・AI 教育強化拠点コンソーシアム」<br/>(<a href="https://www.mi.u-tokyo.ac.jp/6university_consortium.html">https://www.mi.u-tokyo.ac.jp/6university_consortium.html</a>)</li> </ul> </li> <li>● <b>R を用いる場合</b> <ul style="list-style-type: none"> <li>➢ 神戸大学大学院経営学研究科（分寺杏介）「統計的方法論特殊研究（多変量解析）講義資料 3 データの前処理」<br/>(<a href="https://www2.kobe-u.ac.jp/~bunji/files/lecture/MVA/html/chapters/03_data_preprocessing.html">https://www2.kobe-u.ac.jp/~bunji/files/lecture/MVA/html/chapters/03_data_preprocessing.html</a>)</li> </ul> </li> <li>● <b>Python を用いる場合</b> <ul style="list-style-type: none"> <li>➢ 総務省統計局「データサイエンス・オンライン講座 特別セミナー 今日から始められる、Python によるビジネスデータ解析プログラミング講座」<br/>(<a href="https://www.stat.go.jp/dss/pdf/programming/202412teach.pdf">https://www.stat.go.jp/dss/pdf/programming/202412teach.pdf</a>)</li> <li>➢ 九州工業大学 数理・DS・AI 教育推進室「第 5 回 前処理・クレンジング処理」<br/>(<a href="https://kyutech.repo.nii.ac.jp/records/2001414">https://kyutech.repo.nii.ac.jp/records/2001414</a>)</li> </ul> </li> </ul> |
|---|

## ＜基本的なデータクレンジング前後のデータのイメージ＞

### ＜クレンジング前データ＞

クレンジング前データでは、確認ポイントに該当する箇所に表記の誤りやゆれが混在しています。

|  | ID   | 性別 | 都道府県 | 利用媒体 | 世帯人数 | 満足度<br>(10点満点) |                         |
|--|------|----|------|------|------|----------------|-------------------------|
| 表記ゆれの統一<br>「東京都」→<br>「東京都」<br>(全角スペース削除) | 0001 | 男  | 東京都  | PC   | 1    | 8              | 全角・半角の統一<br>「8」→「8」     |
|  | 0002 | 男性 | 東京都  | PCC  | 5    | 2              |                         |
| 性別表記の統一<br>「男性」→「男」                      | 0003 | 男  | 東京   | PC   | 3    | n/a            | 大文字・小文字の統一<br>「pc」→「PC」 |
|  | 0004 | 女  | 東京都  | pc   | 3    | 6              |                         |
| 表記ゆれの統一<br>「東京」→「東京都」                    | 0005 | 男  | 東京都  | PC   | 2    | 7              | 欠損値の統一<br>「99」「-」→「n/a」 |
|  | 0006 | 男  | 東京都  | PC   | 三    | -              |                         |
| 重複データの削除<br>ID「0006」の重複を修正               | 0006 | 女  | 東京都  | PC   | 4    | 99             |                         |

### ＜クレンジング後データ＞

クレンジング後データでは、各項目が統一されています。

| ID   | 性別 | 都道府県 | 利用媒体 | 世帯人数 | 満足度<br>(10点満点) |
|------|----|------|------|------|----------------|
| 0001 | 男  | 東京都  | PC   | 1    | 8              |
| 0002 | 男  | 東京都  | PC   | 5    | 2              |
| 0003 | 男  | 東京都  | PC   | 3    | n/a            |
| 0004 | 女  | 東京都  | PC   | 3    | 6              |
| 0005 | 男  | 東京都  | PC   | 2    | 7              |
| 0006 | 男  | 東京都  | PC   | 3    | n/a            |
| 0007 | 女  | 東京都  | PC   | 4    | n/a            |

- ここまでに示した一時点データの基本的なクレンジングを行うことで、データの利用に向けた準備ができました。**ただちに経年データの作成を行わない場合（将来のデータ利用可能性を考慮した基盤整備や、一時点データの集計や結果の可視化等を主目的とする場合）は、ここまでの作業を経て正確性を高めたデータをご活用ください。**
- 続いて経年データの作成に進む方、作成方法を確認しておきたい方は、2-3 以降にお進み下さい。なお、経年データの作成を進める際は、接続を行うデータセットの数だけ、「基本的なデータクレンジング」を行っておく必要があります。

### ✔ Point 機械判読可能なデータ作成について

- データを活用する第一歩は、PC がそのまま処理できる形（機械が判読可能な形式）でデータを作成・管理することです。調査データが入力されているファイルの形式やレイアウトにも注意を払うことが必要です。
- 人間が見て分かりやすい表でも、PC にとっては非常に扱いづらく、統計ソフトや表計算ソフト（Excel 等）で計算ができず、膨大なデータを一行ずつ手作業で修正するという、非効率な作業が発生してしまうことがあります。
- 機械で判読可能なデータには、いくつかの重要なポイントがあります。主な例として、以下の点が

挙げられます。

- ファイル形式は Excel や CSV など、汎用的な形式であること
- 1 セルにつき 1 つのデータが記録されていること
- 機械判読可能なデータ形式を用いること（日付など）
- 数値データには文字列を含まないこと
- 改行や不要なスペースを使用しないこと
- 「セルの結合」機能を使用しないこと
- これらを含む、具体的なデータ作成ルールについては、以下のガイドラインを参考にしてください。調査票を作る段階から、最終的に出力される「数字の並び（ローデータ）」をイメージしておくことで、その後の分析作業を何倍も効率化することができます。

- 総務省統計局「結果表における機械判読可能なデータ作成に関する表記方法」  
(<https://www.e-stat.go.jp/estat/html/machine-readable-stats-format.pdf>)

## 2-3. 経年データの活用目的を検討する

データの活用目的を検討し、必要に応じて経年データ作成に向かう

目的

### ✔ Point

一時点データの基本的なデータクレンジングを行った後、さらに経年データの作成に進む場合は、まずは**経年データの活用目的を明確にする**必要があります。その目的に沿ってデータクレンジング方針を整理し、効率的に作業を進めましょう。

- 以降では、**経年データの活用の可能性を探りたい**という方に向けて、経年データ作成のためのデータクレンジングの前に検討すべき主な要素について示しています。
- 経年データ作成のためのデータクレンジングは、経年データの活用目的に応じて、以下（1）～（3）の要素を検討することが望まれます。なぜなら、ここでの検討結果が、**使用するデータの範囲や必要となるデータクレンジングの作業量や方針に大きく影響するため**です。

活用目的

- どのような目的で経年データを作成、分析するのか？
- どのような数値の変化を確認したいのか？
- どのような調査項目同士の関係性を見たいのか？
- 分析結果を、どのように用いたいのか？誰に届けたいのか？

|                     |   |
|---------------------|---|
| (1)<br>接続単位         | 個人、世帯、施設、地域など、どの単位でデータをつなげるかを明確にします。                        |
| (2)<br>対象期間         | 直近 3 年間、10 年間、全調査（回答取得）期間など、どの期間のデータを接続対象とするかを明確にします。       |
| (3)<br>データの<br>利用範囲 | 全データを使用するのか、一部の調査項目の回答データのみを抽出して使用するのかなど、接続するデータの範囲を明確にします。 |

- 経年データの活用目的をこの段階で一定程度明確にし、上述の検討要素を事前に整理しておくことで、効率的なデータの取扱いが可能となります。

### 「全国学力・学習状況調査」における検討結果の例

（実際の分析事例については第 5 章参照）

#### 検討結果

|                            |  |
|----------------------------|--|
| 分析目的                       | 経年データを用いて学力が継続的に高い、又は伸びを示す学校や市町村を把握する  |
| 分析・<br>接続単位                | 学校、市町村   |
| 対象期間                       | 平成 31 年～令和 6 年   |
| 分析に<br>使用するデータ<br>セットのイメージ | 分析に使用する一部の調査項目の回答データのみを抽出して使用する。年度ごとに児童生徒単位のデータから学校単位や市町村単位の平均値を算出し、分析に使用するデータセットを作成する |

## 2-4. 経年データの作成に適したデータであるか確認する

### 経年データの作成に適したデータであるか確認する

#### ✓ Point

#### 目的

各種調査等によって取得したデータには、様々な状態のものが存在します。経年データの作成に関する基礎知識を理解した上で、お持ちの調査データが経年データの作成に適しているかチェックリストで確認しましょう。

- 経年データ作成の前提条件として、当然のことながら**複数時点で調査が実施されており（又は実**

施予定であり)、複数回分の調査データがある(又は得られる見込みである)ことが、経年データ作成の前提条件となります。

- しかし、より細かく見ると、①毎回全く同じ個人や事業者等に対して追跡調査を行う場合(いわゆる「パネルデータ」<sup>3</sup>)と、②調査対象は同じであるが、回答する個人等は各時点で異なる場合(いわゆる「繰り返しクロスセクションデータ」<sup>4</sup>)に分けられます。

### <経年データの種類>

#### ①パネルデータ

| 年    | 事業所ID   | 従業員数 | 売上(千円)  |
|------|---------|------|---------|
| 2021 | EST_001 | 80   | 120,663 |
| 2022 | EST_001 | 81   | 121,159 |
| 2023 | EST_001 | 80   | 118,115 |
| 2021 | EST_002 | 45   | 82,601  |
| 2022 | EST_002 | 44   | 90,620  |
| 2023 | EST_002 | 44   | 88,254  |
| 2021 | EST_003 | 68   | 106,262 |
| 2022 | EST_003 | 72   | 108,532 |
| 2023 | EST_003 | 71   | 106,448 |
| 2021 | EST_004 | 49   | 88,095  |
| 2022 | EST_004 | 48   | 89,535  |
| 2023 | EST_004 | 52   | 91,655  |
| 2021 | EST_005 | 44   | 95,430  |
| 2022 | EST_005 | 45   | 102,327 |
| 2023 | EST_005 | 42   | 101,886 |
| 2021 | EST_006 | 44   | 155,560 |
| 2022 | EST_006 | 45   | 159,405 |
| 2023 | EST_006 | 41   | 155,151 |

#### ②繰り返しクロスセクションデータ

| 年    | 児童生徒ID       | 学校ID | 性別 | スコア  |
|------|--------------|------|----|------|
| 2021 | Y21_S001_001 | S001 | 女  | 44.5 |
| 2021 | Y21_S001_002 | S001 | 女  | 61.2 |
| 2021 | Y21_S001_003 | S001 | 男  | 58.9 |
| 2021 | Y21_S001_004 | S001 | 男  | 74   |
| 2021 | Y21_S001_005 | S001 | 女  | 49.4 |
| 2022 | Y22_S001_001 | S001 | 男  | 64   |
| 2022 | Y22_S001_002 | S001 | 女  | 66.5 |
| 2022 | Y22_S001_003 | S001 | 男  | 76.7 |
| 2022 | Y22_S001_004 | S001 | 女  | 60.5 |
| 2022 | Y22_S001_005 | S001 | 男  | 55.6 |
| 2023 | Y23_S001_001 | S001 | 女  | 63.8 |
| 2023 | Y23_S001_002 | S001 | 男  | 77.1 |
| 2023 | Y23_S001_003 | S001 | 女  | 57.6 |
| 2023 | Y23_S001_004 | S001 | 男  | 54.1 |
| 2023 | Y23_S001_005 | S001 | 女  | 67.9 |
| 2021 | Y21_S002_001 | S002 | 女  | 61.4 |
| 2021 | Y21_S002_002 | S002 | 男  | 45.2 |
| 2021 | Y21_S002_003 | S002 | 女  | 60.9 |
| 2021 | Y21_S002_004 | S002 | 女  | 36.7 |
| 2021 | Y21_S002_005 | S002 | 女  | 57.2 |
| 2022 | Y22_S002_001 | S002 | 女  | 78.5 |
| 2022 | Y22_S002_002 | S002 | 女  | 53.7 |
| 2022 | Y22_S002_003 | S002 | 男  | 62   |
| 2022 | Y22_S002_004 | S002 | 男  | 71.9 |
| 2022 | Y22_S002_005 | S002 | 女  | 50.6 |
| 2023 | Y23_S002_001 | S002 | 男  | 75.6 |
| 2023 | Y23_S002_002 | S002 | 男  | 88.2 |
| 2023 | Y23_S002_003 | S002 | 男  | 31.3 |
| 2023 | Y23_S002_004 | S002 | 女  | 39.2 |
| 2023 | Y23_S002_005 | S002 | 女  | 68.8 |

時点ごとに同一の対象者が回答しているのでそのまま接続が可能。

時点毎に異なる対象者が回答しているが、回答者が所属する学校単位にまとめることで接続が可能。

<sup>3</sup> 同じ個体を複数時点で追跡したもの。個体の変化(個体内変化)を追跡することができ、個体固有効果(個体が元々持っている特徴の違い)を取り除くことができるため、因果推論に有効です。

<sup>4</sup> 異なる個体を対象としたクロスセクション(一時点)のデータが複数あるもの。「全国学力・学習状況調査」は毎年小学校第6学年と中学校第3学年を対象に実施されます(=調査対象は同じ)が、解答(回答)している児童生徒は毎年替わっていますので、繰り返しクロスセクションデータに該当します。

- 上図で、①のパネルデータは、各調査時点で同一の事業所が回答しているため、複数時点のデータをそのまま接続し、対象の変化を分析することができます。一方で、②の繰り返しクロスセクションデータは、各調査時点で回答者が異なるため、複数時点のデータをそのまま接続することができません。この場合、**接続が可能な単位にデータを集計・統合することが必要です**<sup>5</sup>。
- このように、**必ずしも同一個人等を追跡した調査である必要はありませんが、同じ調査対象（母集団）に対して実施されている調査であることが**、経年データ作成には不可欠です。
- **各調査時点の調査内容が概ね同じであることも必要**です。定期的に同じ対象に対して実施される調査であっても、毎回調査内容が全く異なる場合は、経年データの作成は難しいと考えられます。
- ただし、調査内容が概ね同じと言っても、実際は、設問文や選択肢などが異なっており、そのままでは接続することが難しいケースが多く見られます。次章において、こうした事象に対するデータクレンジングの方法について詳しく扱います。



### 経年データの活用に注意を要するケース

以下の例の場合は、経年データとしての活用に注意が必要です。

少なくとも時系列での単純な比較は難しいなど、集計結果の読み取りには慎重になる必要があります。

#### ① 難易度調整がされていないテストスコア

- ・ テスト等により測定されるスコアは、複数時点の各テストが同じ難易度であることが担保されていない限り単純な時系列比較を行うことはできません。こうしたテスト間の得点の比較を可能にするため、「項目反応理論（IRT）」を用い難易度調整を行ったテストが作成されることがあります。

#### ② 調査対象の内訳が大きく異なる場合

- ・ 合併等による行政区域の変更や、企業・事業所等の統合などで、調査対象の内実に大きな変更が生じている場合には、時系列の継続性が損なわれる可能性があります。そのため、区域の再集計や系列断絶の明示など、必要な補正を行った上でデータの読み取り、解釈が必要となります。

#### ③ 調査の実施頻度にばらつきがある場合

- ・ 調査が不定期に行われるなど、実施頻度にばらつきがみられる場合は、時系列の変化を読み取る際、調査時点の間隔を踏まえた解釈を行う必要があります。

<sup>5</sup> 全国学力・学習状況調査のデータは、②のクロスセクションデータに該当し、毎年解答（回答）者が異なるため児童生徒単位では接続できません。このため実証的共同研究では、学校単位、市町村単位にデータを集計する（各調査項目の「学校内の平均値」や「市町村内の平均値」を集計するなど）ことで、「W年度のX学校」「Y年度のZ市」という形でデータを接続できるようにしました。



### 経年データの作成に向けたチェックリスト

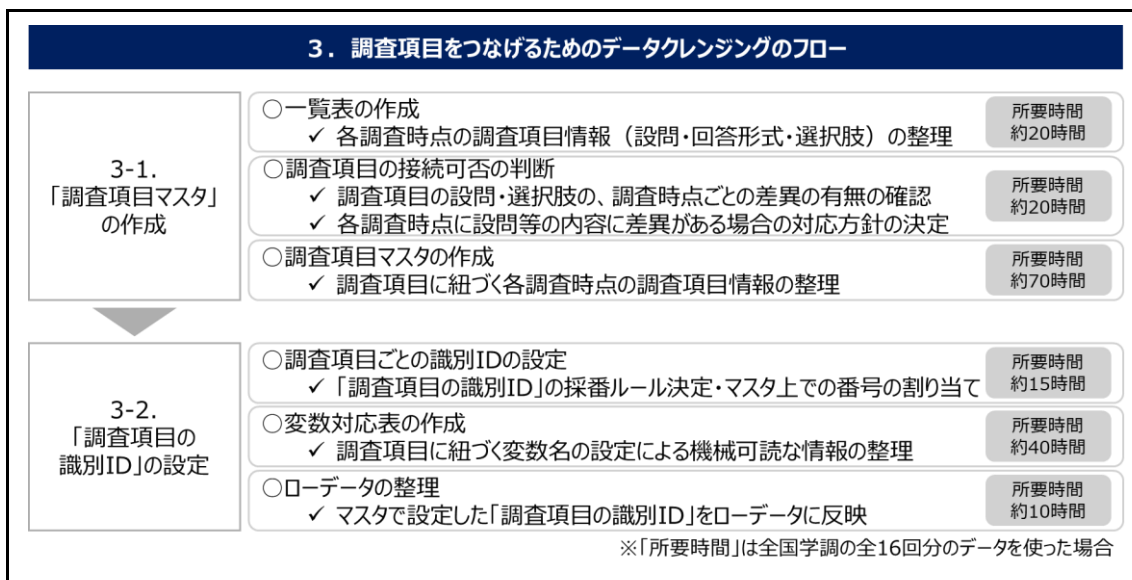
お持ちのデータが経年データの作成に適しているか、確認しましょう。

原則として以下の3つすべてにチェックが入れば、本手引きに沿った経年データの作成が可能です。

- 複数時点分の調査データがある
- 各調査時点において、調査対象が同じである  
(同一個人等への追跡調査である必要はありません)
- 各調査時点において、調査内容が概ね同じである

### 3. 調査項目をつなげるためのデータクレンジング

この章では、調査項目をつなげるためのデータクレンジングに焦点を当て、複数時点のデータを調査項目ごとに接続し、各調査項目が複数時点で変化している場合、それを可視化する具体的な作業手順について説明します。



#### 3-1. 「調査項目マスタ」を作成する

##### 目的

複数時点の調査票の設問文や回答の選択肢などを一覧化してまとめた「調査項目マスタ」を作成する

##### Point

複数時点のデータをつなげるには、**接続対象とする期間の中で各調査項目情報の意味や構造が一致しているか**を確認する必要があります。しかし、実際には設問文・回答形式・選択肢が調査時点ごとに異なることが多いため、まず調査項目マスタを作成し、項目の対応関係を確認できるようにします。

##### 3-1-1. 手順①：複数時点の調査項目情報の一覧表を Excel で作成する

- 調査票をもとに各時点の設問文・回答形式・選択肢を整理し、一覧表を作成しましょう。なお、ここで一覧化するものは、データの中身（回答値）ではないことにご留意ください。

## 一覧表に含める情報

- 一覧表には、以下の情報を含めます。

|                  |   |
|------------------|---|
| 整理番号             | 調査項目名ごとに付与する固有番号<br>調査項目マスタの作成（手順 3-1-2）で使用する           |
| 調査項目番号<br>（設問番号） | 各時点の調査票における大問・枝問番号                                      |
| 調査項目名<br>（設問文）   | 調査票の調査項目名（アンケート調査の場合は設問文）                               |
| 回答の選択肢           | 調査票の選択肢番号とその内容<br>※数値回答の場合、選択肢がないため、選択肢「（数値回答）」と表記する    |
| 回答形式             | 各調査項目の回答形式<br>単一回答（SA）、複数回答（MA）、数値回答（NA/NUM）、自由回答（FA）／等 |

## 一覧表の見方

- ① 縦軸は各設問が、横軸は各時点の調査項目情報（調査項目番号、調査項目名、回答形式）が入ります。
- ② 最も古い時点の調査項目情報が一番左に、最も新しい調査項目情報が一番右にあります。調査項目名や選択肢の内容がどのように変化しているかが一覧で分かる構造になっています。

### 一覧表の例（※全国学力・学習状況調査の調査項目を使用）

| 整理<br>番号 | H19              |                     |  |          | H20              |                      |  |          | R5               |  |  |          | R6               |  |  |          |
|----------|------------------|---------------------|--|----------|------------------|----------------------|--|----------|------------------|--|--|----------|------------------|--|--|----------|
|          | 調査項目番号<br>(設問番号) | 調査項目名<br>(設問文)      | 選択肢  | 回答<br>形式 | 調査項目番号<br>(設問番号) | 調査項目名<br>(設問文)       | 選択肢  | 回答<br>形式 | 調査項目番号<br>(設問番号) | 調査項目名<br>(設問文)                                   | 選択肢  | 回答<br>形式 | 調査項目番号<br>(設問番号) | 調査項目名<br>(設問文)                                     | 選択肢  | 回答<br>形式 |
| 1        | 1                | 朝食を毎日食べていますか        | ①している<br>②どちらかといえば、<br>している<br>③あまりしていない<br>④全くしていない | SA       | 1                | 朝食を毎日食べていますか         | ①している<br>②どちらかといえば、<br>している<br>③あまりしていない<br>④全くしていない | SA       | 1                | 朝食を毎日食べていますか                                     | ①している<br>②どちらかといえば、<br>している<br>③あまりしていない<br>④全くしていない               | SA       | 1                | 朝食を毎日食べていますか                                       | ①している<br>②どちらかといえば、<br>している<br>③あまりしていない<br>④全くしていない               | SA       |
| 2        | 4                | 毎日、同じくらいの時刻に寝ていますか  | ①している<br>②どちらかといえば、<br>している<br>③あまりしていない<br>④全くしていない | SA       | 3                | 毎日、同じくらいの時刻に寝ていますか   | ①している<br>②どちらかといえば、<br>している<br>③あまりしていない<br>④全くしていない | SA       | 2                | 毎日、同じくらいの時刻に寝ていますか                               | ①している<br>②どちらかといえば、<br>している<br>③あまりしていない<br>④全くしていない               | SA       | 2                | 毎日、同じくらいの時刻に寝ていますか                                 | ①している<br>②どちらかといえば、<br>している<br>③あまりしていない<br>④全くしていない               | SA       |
| 3        | 5                | 毎日、同じくらいの時刻に起きていますか | ①している<br>②どちらかといえば、<br>している<br>③あまりしていない<br>④全くしていない | SA       | 4                | 毎日、同じくらいの時刻に起きていますか  | ①している<br>②どちらかといえば、<br>している<br>③あまりしていない<br>④全くしていない | SA       | 3                | 毎日、同じくらいの時刻に起きていますか                              | ①している<br>②どちらかといえば、<br>している<br>③あまりしていない<br>④全くしていない               | SA       | 3                | 毎日、同じくらいの時刻に起きていますか                                | ①している<br>②どちらかといえば、<br>している<br>③あまりしていない<br>④全くしていない               | SA       |
| 4        | -                | -                   | -  | -        | -                | -                    | -  | -        | 4                | 授業で学んだことを、次の学習や実生活に結びつけて考えたり、生かしたりすることができると思いますか | ①当てはまる<br>②どちらかといえば、<br>当てはまる<br>③どちらかといえば、<br>当てはまらない<br>④当てはまらない | SA       | 4                | 授業で学んだことを、次の学習や実生活に結びつけて考えたり、生かしたり、生かすことができると思いますか | ①当てはまる<br>②どちらかといえば、<br>当てはまる<br>③どちらかといえば、<br>当てはまらない<br>④当てはまらない | SA       |
| 5        | -                | -                   | -  | -        | -                | -                    | -  | -        | -                | -  | -  | -        | 5                | 分からないことや詳しく知りたいことがあったときに、自分で学び方を考え、工夫することばできますか    | ①できている<br>②どちらかといえば、<br>できている<br>③どちらかといえば、<br>できていない<br>④できていない   | SA       |
| 212      | -                | -                   | -  | -        | 25               | 家で自分で計画を立てて勉強をしていますか | ①している<br>②どちらかといえば、<br>している<br>③あまりしていない<br>④全くしていない | -        | 16               | 家で自分で計画を立てて勉強をしていますか（学校の授業の予習や復習を含みます）           | ①よくしている<br>②ときどきしている<br>③あまりしていない<br>④全くしていない                      | SA       | -                | -  | -  | -        |

## 一覧表の作成手順

- ① 接続対象とする期間の全ての調査票を準備しましょう。
- ② 作成する一覧表の一番左の列に整理番号を 1 から順に振ります。

|    | A    | B | C | D | E | F | G |
|----|------|---|---|---|---|---|---|
|    | 整理番号 |   |   |   |   |   |   |
| 1  |      |   |   |   |   |   |   |
| 2  | 1    |   |   |   |   |   |   |
| 3  | 2    |   |   |   |   |   |   |
| 4  | 3    |   |   |   |   |   |   |
| 5  | 4    |   |   |   |   |   |   |
| 6  | 5    |   |   |   |   |   |   |
| 7  | 6    |   |   |   |   |   |   |
| 8  | 7    |   |   |   |   |   |   |
| 9  | 8    |   |   |   |   |   |   |
| 10 | 9    |   |   |   |   |   |   |
| 11 | 10   |   |   |   |   |   |   |
| 12 | ⋮    |   |   |   |   |   |   |

- ③ 最新時点の調査票を確認し、整理番号 1 の行に調査項目番号 1（アンケート調査の場合は「設問 1」に該当）の情報（調査項目番号（設問番号）、調査項目名（設問文）、選択肢、回答形式）を入力し、次の整理番号 2 の行に調査項目番号 2（設問 2）の情報・・・と上から順に、各行に調査項目情報を入力します。

|   | A    | B                | C  | D  | E    | F | G | H | I | J |
|---|------|------------------|--|--|------|---|---|---|---|---|
| 1 | 整理番号 | 調査項目番号<br>(設問番号) | 調査項目名<br>(設問文)                                     | 選択肢  | 回答形式 |   |   |   |   |   |
| 2 | 1    | 1                | 朝食を毎日食べていますか                                       | ①している<br>②どちらかといえば、している<br>③あまりしていない<br>④全くしていない           | SA   |   |   |   |   |   |
| 3 | 2    | 2                | 毎日、同じくらいの時刻に寝ていますか                                 | ①している<br>②どちらかといえば、している<br>③あまりしていない<br>④全くしていない           | SA   |   |   |   |   |   |
| 4 | 3    | 3                | 毎日、同じくらいの時刻に起きていますか                                | ①している<br>②どちらかといえば、している<br>③あまりしていない<br>④全くしていない           | SA   |   |   |   |   |   |
| 5 | 4    | 4                | 授業で学んだことを、次の学習や実生活に結びつけて考えたり、生かしたりすることができるとおもいますか  | ①当てはまる<br>②どちらかといえば、当てはまる<br>③どちらかといえば、当てはまらない<br>④当てはまらない | SA   |   |   |   |   |   |
| 6 | 5    | 5                | 分からないことや詳しく知りたいことがあったときに、自分で学び方を考え、工夫することはできていますか。 | ①できている<br>②どちらかといえば、できている<br>③どちらかといえば、できていない<br>④できていない   | SA   |   |   |   |   |   |
| 7 | ⋮    | ⋮                | ⋮  | ⋮  | ⋮    |   |   |   |   |   |
| 8 |      |                  |  |  |      |   |   |   |   |   |
| 9 |      |                  |  |  |      |   |   |   |   |   |

④ 最新時点の調査項目情報を入力した4列の左隣に新しく4列を挿入し、最新時点から1期古い時点の設問の調査項目情報を入力します。

- この際、最新時点と設問項目番号や調査項目情報が異なっていることがあるため、最新時点の一覧表と当該調査票を見比べ、適切な行に調査項目情報を入力していきます。
- 具体的には、まずは各時点で調査項目情報が変更されていないという前提に立ち、最新年度の同じ調査項目番号の設問と見比べ、調査項目名や選択肢が同内容又は概ね似た内容であれば同じ行に入力します。
- **同じ調査項目番号の調査項目名や選択肢と内容が全く異なる場合**、他の調査項目番号の調査項目名や選択肢を確認し、同内容又は概ね似た内容がある場合は、その行に入力します。
- **最新時点のいずれの調査項目情報にも同内容のものがない場合**（当調査項目が最新時点で廃止されている場合又は調査項目名等に明らかな変更が生じている場合）、一覧表の一番下の行に当調査項目情報を新しく追記しましょう。この際、最新時点の調査項目情報には「-」を入力します。
- **調査項目情報が最新時点で新設されており、入力時点の調査票には設定されていない場合**、最新時点より前の調査項目情報には「-」を入力します。
- この際、同じ行に記載すべきか、異なる行に記載すべきか悩む調査項目情報については、3-1-2の「調査項目マスタの作成手順」の手順④で判断を行うため、**この段階では明らかに異なる行とすべきものについてのみ行を変更するようにしましょう。**

| R5   |                  |  |  | R6   |                  |  |  |      |
|------|------------------|--|--|------|------------------|--|--|------|
| 整理番号 | 調査項目番号<br>(設問番号) | 調査項目名<br>(設問文)                         | 選択肢  | 回答形式 | 調査項目番号<br>(設問番号) | 調査項目名<br>(設問文)                                     | 選択肢  | 回答形式 |
| 1    | 1                | 朝食を毎日食べていますか                           | ①している<br>②どちらかといえば、している<br>③あまりしていない<br>④全くしていない     | SA   | 1                | 朝食を毎日食べていますか                                       | ①している<br>②どちらかといえば、している<br>③あまりしていない<br>④全くしていない     |      |
| 2    | 2                | 毎日、同じくらいの時刻に寝ていますか                     | ①している<br>②どちらかといえば、している<br>③あまりしていない<br>④全くしていない     | SA   | 2                | 毎日、同じくらいの時刻に寝ていますか                                 | ①している<br>②どちらかといえば、している<br>③あまりしていない<br>④全くしていない     |      |
| 3    | 3                | 毎日、同じくらいの時刻に起きていますか                    | ①している<br>②どちらかといえば、している<br>③あまりしていない<br>④全くしていない     | SA   | 3                | 毎日、同じくらいの時刻に起きていますか                                | ①している<br>②どちらかといえば、している<br>③あまりしていない<br>④全くしていない     |      |
| 4    | 4                | 授業で学んだことを、ほかの学習や趣味の生活などで生かしていますか       | ①当てはまる<br>②どちらかといえば、当てはまる<br>③あまり当てはまらない<br>④当てはまらない | SA   | 4                | 授業で学んだことを、次の学習や実生活に結びつけて考えたり、生かしたりすることができると感じていますか | ①当てはまる<br>②どちらかといえば、当てはまる<br>③あまり当てはまらない<br>④当てはまらない |      |
| 5    | -                | -                                      | -  | -    | 5                | 自分ですべての学習や実生活に責任を持って取り組むことができると感じていますか             | ①できている<br>②どちらかといえば、できている<br>③あまりできていない<br>④できていない   |      |
| 212  | 16               | 家で自分で計画を立てて勉強をしていますか（学校の授業の予習や復習を指します） | ①よくしている<br>②ときどきしている<br>③あまりしていない<br>④全くしていない        | SA   | -                | -  | -  |      |

入力した調査時点の左に列を追加し、1期古い調査時点の調査項目情報を整理番号の列の右側に入力

⑤ 同様に、④で入力した時点から1期古い時点の調査項目情報を入力する…を繰り返し、接続を行う調査のうち一番古い時点の調査項目情報まで入力します。

- ここでは作業④を一番古い時点まで繰り返し行うことになります。
- この際、最新時点から2期古い時点の調査項目情報を入力する場合は、1期古い時点の調査項目情報と最新時点の情報も含めて照合します。

3期古い時点の場合も同様に、2期古い時点、1期古い時点、最新時点の調査項目情報を合わせて照合します。

つまり、入力する時点から1期新しい時点の調査項目情報を基準としつつ、それまでに入力した全ての時点の調査項目情報と見比べる必要があります。

| 整理番号 | H19              |                     |  |      | H20              |                      |  |      | R6               |  |  |      |
|------|------------------|---------------------|--|------|------------------|----------------------|--|------|------------------|--|--|------|
|      | 調査項目番号<br>(設問番号) | 調査項目名<br>(設問文)      | 選択肢  | 回答形式 | 調査項目番号<br>(設問番号) | 調査項目名<br>(設問文)       | 選択肢  | 回答形式 | 調査項目番号<br>(設問番号) | 調査項目名<br>(設問文)                                     | 選択肢  | 回答形式 |
| 1    | 1                | 朝食を毎日食べていますか        | ① している<br>② どちらかといえば、している<br>③ あまりしていない<br>④ 全くしていない | SA   | 1                | 朝食を毎日食べていますか         | ① している<br>② どちらかといえば、している<br>③ あまりしていない<br>④ 全くしていない | SA   | 1                | 朝食を毎日食べていますか                                       | ① している<br>② どちらかといえば、している<br>③ あまりしていない<br>④ 全くしていない         | SA   |
| 2    | 4                | 毎日、同じくらいの頻度に行っていますか | ① している<br>② どちらかといえば、している<br>③ あまりしていない<br>④ 全くしていない | SA   | 3                | 毎日、同じくらいの頻度に行っていますか  | ① している<br>② どちらかといえば、している<br>③ あまりしていない<br>④ 全くしていない | SA   | 2                | 毎日、同じくらいの頻度に行っていますか                                | ① している<br>② どちらかといえば、している<br>③ あまりしていない<br>④ 全くしていない         | SA   |
| 3    | 5                | 毎日、同じくらいの頻度に行っていますか | ① している<br>② どちらかといえば、している<br>③ あまりしていない<br>④ 全くしていない | SA   | 4                | 毎日、同じくらいの頻度に行っていますか  | ① している<br>② どちらかといえば、している<br>③ あまりしていない<br>④ 全くしていない | SA   | 3                | 毎日、同じくらいの頻度に行っていますか                                | ① している<br>② どちらかといえば、している<br>③ あまりしていない<br>④ 全くしていない         | SA   |
| 4    | -                | -                   | -  | -    | -                | -                    | -  | -    | 4                | 授業で学んだことを、次の学習や実生活に結びつけて考えたり、生かしたりすることがあると思いますか    | ① 当てはまる<br>② どちらかといえば、当てはまる<br>③ 当てはまらない<br>④ 当てはまらない        | SA   |
| 5    | -                | -                   | -  | -    | -                | -                    | -  | -    | 5                | 分からないことや興味・関心のあることがあったときに、自分で学び方を考え、工夫することを行っていますか | ① できている<br>② どちらかといえば、できている<br>③ どちらかといえば、できていない<br>④ できていない | SA   |
| 212  | -                | -                   | -  | -    | 25               | 朝で自分で計画を立てて勉強をしていますか | ① している<br>② どちらかといえば、している<br>③ あまりしていない<br>④ 全くしていない | -    | -                | -  | -  | -    |

手順④を繰り返し、接続対象としている最も古い調査時点の調査項目情報まで入力する

⑥ 最後に、表内一番左の列の整理番号を確認します。手順④で行が追加された場合、追加行には整理番号が振られていませんので、追加で入力します。

| 整理番号 | H19              |                |       |      | R6               |                |       |      |
|------|------------------|----------------|-------|------|------------------|----------------|-------|------|
|      | 調査項目番号<br>(設問番号) | 調査項目名<br>(設問文) | 選択肢   | 回答形式 | 調査項目番号<br>(設問番号) | 調査項目名<br>(設問文) | 選択肢   | 回答形式 |
| 1    | ●                | ●●●●           | ●●●●  | ●    | ●                | ●●●●           | ●●●●  | ●    |
| 2    | ×                | ×××××          | ××××  | ×    | ×                | ×××××          | ××××  | ×    |
| 3    | ▲                | ▲▲▲▲           | ▲▲▲   | ▲    | ▲                | ▲▲▲▲           | ▲▲▲   | ▲    |
| 4    | ～                | ～～～～           | ～～～   | ～    | ～                | ～～～～           | ～～～   | ～    |
| 5    | ■                | ■ ■ ■ ■        | ■ ■ ■ | ■    | ■                | ■ ■ ■ ■        | ■ ■ ■ | ■    |
| 212  | ○                | ○○○○           | ○○○   | ○    | -                | -              | -     | -    |
| 213  | ×                | ×××××          | ××××  | ×    | -                | -              | -     | -    |
| 214  | △                | △△△△           | △△△   | △    | -                | -              | -     | -    |
| 215  | ～                | ～～～～           | ～～～   | ～    | -                | -              | -     | -    |
| 216  | □                | □□□□           | □□□   | □    | -                | -              | -     | -    |

最新時点で採用されている調査項目

最新時点では廃止されている調査項目

整理番号が入力されていない行に整理番号を追加する

### 3-1-2.手順②：複数時点の調査項目情報の一覧表から「調査項目マスタ」を作成する

- ここでは、3-1-1 で作成した複数時点の調査項目情報の一覧表から、時点間での調査項目名（設問文）や選択肢の変更を整理した「**調査項目マスタ**」を作成します。後に、このマスタで整理した対応方針に沿って、調査項目を接続する作業を行います。

#### ✔ Point 調査項目マスタとは

「調査項目マスタ」とは、**接続対象とする期間における、各調査項目の調査項目名（設問文）・選択肢等の変更を整理した表**です。基準となる時点の調査項目をベースに、調査項目名（設問文）や選択肢等の調査項目情報にどのように変更が生じているのか把握することができ、経年でのバージョン管理が容易になります。

- このマスタは、後に第4章でデータを接続する際に、同じ調査項目であるかどうかを判断する材料となります。もし途中で調査項目名（設問文）や選択肢に変更があり、別の調査項目として扱う必要がある場合は、行を分けて管理します。

#### 調査項目マスタに含める情報

- 調査項目マスタには、以下の情報を含めます。

|                  |   |
|------------------|---|
| 整理番号             | 調査項目ごとに付与する固有番号   |
| バージョン            | 同じ調査項目であっても、調査項目名や選択肢の変更により、異なるバージョンとして扱う必要がある場合に更新する番号                             |
| 調査項目名<br>（設問文）   | 調査票の調査項目名（アンケート調査の場合は設問文）<br>（経年接続時の基準となる調査項目名（設問文）。誤脱や軽微な変更がある場合は、修正した最適なものを記載する。） |
| 回答の選択肢           | 調査票の選択肢番号とその内容  |
| 回答形式             | 各調査項目の回答形式<br>単一回答（SA）、複数回答（MA）、数値回答（NA/NUM）、自由回答（FA）／等                             |
| 調査項目番号<br>（設問番号） | 各時点の調査票における大問・枝問番号<br>（当該調査項目のデータが取得されていない年には「-（ハイフン）」が入り、未取得であることがわかる。）            |

## 調査項目マスタの見方

- ① 縦軸には各調査項目が、横軸には調査項目情報が入ります。
  - 横軸について、一覧表では時点ごとに調査項目情報がまとっていましたが、調査項目マスタでは、調査項目を接続する際に統一基準となる調査項目をベースとした情報が記載されています。
- ② 各調査項目でいつどのようなバージョン変更があったのかが一目で分かるような構造になっています。
- ③ 調査項目マスタを作成後、新たに調査を実施する際、調査項目を追加する場合や、調査項目名や選択肢を変更する場合は、調査項目マスタに追記して、調査項目の変更履歴を残すようにしましょう。

調査項目マスタの例 ※全国学力・学習状況調査の調査項目を使用

| 整理<br>番号 | バー<br>ジョン | 調査項目名<br>(設問文)                                      | 選択肢  | 回答<br>形式 | 調査項目番号<br>(設問番号) |     |     |     | 備考欄                    |
|----------|-----------|---|--|----------|------------------|-----|-----|-----|------------------------|
|          |           |   |  |          | H19              | H20 | ... | R6  |                        |
| 1        | 1         | 朝食を毎日食べていますか  | ①している<br>②どちらかといえば、している<br>③あまりしていない<br>④全くしていない   | SA       | 1                | 1   | ... | 1   |                        |
| ...      | ...       | ...   | ...  | ...      | ...              | ... | ... | ... |                        |
| 108      | 1         | 地域や社会をよくするために何をすべきかを考えることがありますか                     | ①当てはまる<br>②どちらかといえば、当てはまる<br>③どちらかといえば、当てはまらない<br>④当てはまらない                                     | SA       | 29               | 30  | ... | -   |                        |
| 108      | 2         | 地域や社会をよくするために何かしてみたいと思いますか                          | ①当てはまる<br>②どちらかといえば、当てはまる<br>③どちらかといえば、当てはまらない<br>④当てはまらない                                     | SA       | -                | -   | ... | 25  |                        |
| ...      | ...       | ...   | ...  | ...      | ...              | ... | ... | ... |                        |
| 137      | 1         | 携帯電話やスマートフォンで通話やメールをしていますか                          | ①携帯電話を持っていない<br>②全く、または、ほとんどしていない<br>③ときどきしている<br>④ほぼ毎日している                                    | SA       | 19               | 19  | ... | -   | H19からH20にかけて調査項目名に変更あり |
| ...      | ...       | ...   | ...  | ...      | ...              | ... | ... | ... |                        |
| 180      | 1         | 普段(月曜日から金曜日)、1日当たりどれくらいの時間、テレビやビデオ・DVDを見たり、聞いたりしますか | ①全く見たり、聞いたりしない<br>②1時間より少ない<br>③1時間以上、2時間より少ない<br>④2時間以上、3時間より少ない<br>⑤3時間以上、4時間より少ない<br>⑥4時間以上 | SA       | 17               | 12  | ... | -   | H28からH29にかけて選択肢に変更あり   |
| ...      | ...       | ...   | ...  | ...      | ...              | ... | ... | ... |                        |
| 300      | 1         | 友達に会うのは楽しいと思いますか                                    | ①そう思う<br>②どちらかといえば、そう思う<br>③どちらかといえば、そう思わない<br>④そう思わない   | SA       | 36               | 29  | ... | -   |                        |

※ SA : 単一回答

## 調査項目マスタの作成手順

### ① 整理番号の記入

- 表の一番左の列に複数時点の調査項目情報の一覧表に記載の整理番号をコピー＆ペーストしましょう。

|    | A    | B | C | D | E | F | G |
|----|------|---|---|---|---|---|---|
| 1  | 整理番号 |   |   |   |   |   |   |
| 2  | 1    |   |   |   |   |   |   |
| 3  | 2    |   |   |   |   |   |   |
| 4  | 3    |   |   |   |   |   |   |
| 5  | 4    |   |   |   |   |   |   |
| 6  | 5    |   |   |   |   |   |   |
| 7  | 6    |   |   |   |   |   |   |
| 8  | 7    |   |   |   |   |   |   |
| 9  | 8    |   |   |   |   |   |   |
| 10 | 9    |   |   |   |   |   |   |
| 11 | 10   |   |   |   |   |   |   |
| 12 | ⋮    |   |   |   |   |   |   |

一覧表から  
コピー＆ペーストする

### ② バージョン列を追加

- 調査項目マスタでは、同じ調査項目でも異なるバージョンとして扱うものは行を分けて管理するため、バージョン番号を記入できる列を作成します。複数時点の調査項目情報の一覧表で整理されている調査項目は、それぞれバージョン 1 から数えるため、①の右隣の列のすべての行に「1」と記入します。

|    | A        | B     | C | D | E | F | G |
|----|----------|-------|---|---|---|---|---|
| 1  | 整理<br>番号 | バージョン |   |   |   |   |   |
| 2  | 1        | 1     |   |   |   |   |   |
| 3  | 2        | 1     |   |   |   |   |   |
| 4  | 3        | 1     |   |   |   |   |   |
| 5  | 4        | 1     |   |   |   |   |   |
| 6  | 5        | 1     |   |   |   |   |   |
| 7  | 6        | 1     |   |   |   |   |   |
| 8  | 7        | 1     |   |   |   |   |   |
| 9  | 8        | 1     |   |   |   |   |   |
| 10 | 9        | 1     |   |   |   |   |   |
| 11 | 10       | 1     |   |   |   |   |   |
| 12 | ⋮        | ⋮     |   |   |   |   |   |

整理番号を振った  
すべての行に「1」を記入

### ③ 調査項目名（設問文）の記入

- ②の右隣の列に、一覧表から、その調査項目の統一基準となる調査項目名（設問文）をコピー＆ペーストしましょう。統一基準とする調査項目名（設問文）は、後に示す「調査項目マスタ作成時のポイント」（p.39）に沿って、誤脱等のない適切なものを選択しましょう。必ずしも最新時点の調査項目名（設問文）に統一する必要はありません。

|   | A        | B     | C   | D | E | F |
|---|----------|-------|---|---|---|---|
| 1 | 整理<br>番号 | バージョン | 調査項目名<br>(設問文)                                    |   |   |   |
| 2 | 1        | 1     | 朝食を毎日食べていますか                                      |   |   |   |
| 3 | 2        | 1     | 毎日、同じくらいの時刻に寝ていますか                                |   |   |   |
| 4 | 3        | 1     | 毎日、同じくらいの時刻に起きていますか                               |   |   |   |
| 5 | 4        | 1     | 授業で学んだことを、次の学習や実生活に結び付けて考えたり、生かしたりすることができると思いますか  |   |   |   |
| 6 | 5        | 1     | 分からないことや詳しく知りたいことがあったときに、自分で学び方を考え、工夫することはできていますか |   |   |   |
| 7 | ⋮        | ⋮     |   |   |   |   |

統一基準となる設問文を  
一覧表から  
コピー＆ペーストする

### ④ 調査項目名（設問文）の差異による、調査項目のバージョン分け

- ③の作業の中で、同じ調査項目であっても一部の単語が途中から変更・追加されているなど、差異が生じているものがある場合は、全く同じものとして接続するか、同じ調査項目の新しいバージョンとして管理するかを判断します。
- 後に示す「調査項目マスタ作成時のポイント」（p.39）に沿って、調査項目名（設問文）

に差異が生じているが、同じバージョンとして接続するとした調査項目については、調査項目名（設問文）の右隣に備考欄の列を設け、いつ差異が生じているかわかるようにします。

| A    | B     | C                                       | D  |
|------|-------|---|--|
| 整理番号 | バージョン | 調査項目名<br>(設問文)                          | 備考欄  |
| 1    | 1     | 朝食を毎日食べていますか                            |  |
| 2    | 1     | 毎日、同じくらいの時刻に寝ていますか                      |  |
| ⋮    | ⋮     | ⋮                                       |  |
| 22   | 1     | 携帯電話やスマートフォンの使い方について、家の人と約束したことを守っていますか | H24からH25にかけて調査項目名に変更あり   |
| 23   | 1     | 体か間                                     | 例) H24からH25にかけて差異が生じているが、基準となる調査項目名に統一して経年接続する場合、備考欄に示し、変更があることがわかるようにしておく |
| ⋮    | ⋮     | ⋮                                       |  |

- 調査項目名（設問文）の差異により、調査項目の趣旨が異なると判断した場合は、当該調査項目の1つ下に行を追加し、同じ調査項目の新バージョンとして管理します。
- 整理番号は元の調査項目と同じ番号、バージョン番号は「2」と記入し、調査項目名（設問文）は新しいバージョンにおいて統一基準となるものを記入します。
- 同じ調査項目でバージョン番号が3、4…と続く場合は、それに応じて行を追加し、同様の作業を行います。

| A    | B     | C                               | D  |
|------|-------|---------------------------------|--|
| 整理番号 | バージョン | 調査項目名<br>(設問文)                  | 備考欄  |
| ⋮    | ⋮     | ⋮                               |  |
| 107  | 1     | 新聞を読んでいますか                      | 同じ調査項目でも趣旨が異なると判断し、経年接続しない場合は、 <u>バージョン2</u> として、元の調査項目の一段下に記入する |
| 108  | 1     | 地域や社会をよくするために何をすべきかを考えることがありますか |  |
| 108  | 2     | 地域や社会をよくするために何かしてみたいと思いますか      |  |
| 109  | 1     | 放課後や週末に何をして過ごすことが多いですか          |  |
| ⋮    | ⋮     | ⋮                               |  |

## ⑤ 選択肢の記入

- 調査項目名（設問文）の右隣に1列追加し、複数時点の調査項目情報の一覧表から、その調査項目の統一基準となる選択肢をコピー & ペーストしましょう。基準とする選択肢は、基準とした調査項目名（設問文）と対応するものとします。複数のバージョンがある調査項目には、それぞれのバージョンについて選択肢を記入しましょう。バージョンが異なっても選択肢は同じ場合もあれば、バージョンの変更に応じて選択肢も変更されている場合もあります。

| 整理番号 | バージョン | 調査項目名<br>(設問文)     | 選択肢  | 備考欄                  |
|------|-------|--------------------|--|----------------------|
| 1    | 1     | 朝食を毎日食べていますか       | ①している<br>②どちらかといえば、している<br>③あまりしていない<br>④全くしていない           |                      |
| 2    | 1     | 毎日、同じくらいの時刻に寝ていますか | ①している<br>②どちらかといえば、している<br>③あまりしていない<br>④全くしていない           |                      |
| 3    | 1     | 毎日、同じくらいの時刻に起きますか  | ①している<br>②どちらかといえば、している<br>③あまりしていない<br>④全くしていない           |                      |
| 4    |       |                    | ①当てはまる<br>②どちらかといえば、当てはまる<br>③どちらかといえば、当てはまらない<br>④当てはまらない | R3からR4にかけて調査項目名に変更あり |
| 5    |       |                    | ①できている<br>②どちらかといえば、できている<br>③どちらかといえば、できていない<br>④できていない   |                      |

調査項目名と備考欄  
の間に、統一基準となる  
選択肢を一覧表から  
コピー & ペーストする

#### ⑥ 選択肢の差異による、調査項目のバージョン分け

- ⑤の作業の中で、同じ調査項目であっても、選択肢に変更・追加があるなど差異が生じているものがある場合は、全く同じものとして接続するか、同じ調査項目の新しいバージョンとして管理するかを判断します。
- 後に示す「調査項目マスタ作成時のポイント」(p.39)に沿って、選択肢の表現が異なるが、回答内容に影響しない場合や、選択肢の順番が入れ替わっているが、内容は同じ場合等、同じバージョンとして接続すると判断した場合は、備考欄に示し、いつ差異が生じているか分かるようにしておきましょう。

| 整理番号 | バージョン | 調査項目名<br>(設問文)                                      | 選択肢  | 備考欄  |
|------|-------|---|--|--|
| 1    | 1     | 朝食を毎日食べていますか  | ①している<br>②どちらかといえば、している<br>③あまりしていない<br>④全くしていない   |  |
| ...  | ...   | ...   | ...  | ...  |
| 180  | 1     | 普段(月曜日から金曜日)、1日当たりどれくらいの時間、テレビやビデオ・DVDを見たり、聞いたりしますか | ①全く見たり、聞いたりしない<br>②1時間より少ない<br>③1時間以上、2時間より少ない<br>④2時間以上、3時間より少ない<br>⑤3時間以上、4時間より少ない<br>⑥4時間以上 | H28からH29にかけて選択肢に変更あり                             |
| ...  | ...   | ...   | ...  | ...  |
| 300  |       |   | 思う<br>らかといえば、そう思う<br>らかといえば、そう思わない<br>思わない   | H15からH16にかけて調査項目名に変更あり<br>H23からH24にかけて調査項目名に変更あり |
| ...  | ...   | ...   | ...  | ...  |

選択肢に差異があるが、基準となる選択肢に統一して経年接続する場合、備考欄に示し、変更があることがわかるようにしておく

- 異なる調査項目として扱うと判断した場合は、④の作業と同様に、当該調査項目の1つ下に行を追加し、同じ調査項目の異なるバージョンとして管理します。
- 整理番号は元の調査項目と同じ番号、バージョン番号は「2」と記入し、調査項目名（設問文）は元の調査項目名をコピー＆ペーストします。選択肢は新しいバージョンにおいて統一基準となるものを記入します。
- 同じ調査項目でバージョン番号が3、4…と続く場合は、それに応じて行を追加し、同様の作業を行います。

| 整理番号 | バージョン | 調査項目名<br>(設問文)          | 選択肢  | 備考欄  |
|------|-------|-------------------------|--|--|
| ...  | ...   | ...                     | ...  | ...  |
| 122  | 1     | 土曜日の午後は、何をして過ごすことが多いですか | ①学習塾など学校や家以外の場所で勉強している<br>②習い事やスポーツ、地域の活動に参加している<br>③家で勉強や読書をしている<br>④家でテレビやビデオ・DVDを見たり、ゲームをしたりしている<br>⑤家族と過ごしている<br>⑥友達と遊んでいる<br>⑦1～6以外   |  |
| 122  | 2     | 土曜日の午後は、何をして過ごすことが多いですか | ①学校で授業を受けている<br>②家で勉強や読書をしている<br>③学習塾など学校や家以外の場所で勉強している<br>④習い事（スポーツに関する習い事を除く）をしている<br>⑤スポーツ（スポーツに関する習い事を含む）をしている<br>⑥地域の活動に参加している<br>⑦家でテレビやビデオ・DVDを見たり、ゲームをしたり、インターネットをしたりしている<br>⑧家族と過ごしている<br>⑨友達と遊んでいる | 選択肢に差異があり経年接続しない場合は、バージョン2として元の調査項目の一段下に記入する |
| ...  | ...   | ...                     | ...  | ...  |

⑦ 回答形式の記入

- 選択枝の右隣の列に1列追加し、一覧表から、調査項目の回答形式をコピー＆ペーストしましょう。

| 整理番号 | バージョン | 調査項目名<br>(設問文)                                      | 選択枝  | 回答形式 | 備考欄  |
|------|-------|---|--|------|--|
| 1    | 1     | 朝食を毎日食べていますか  | ①している<br>②どちらかといえば、している<br>③あまりしていない<br>④全くしていない                                     | SA   |  |
| ...  | ...   | ...   | ...  | ...  | ...  |
| 180  | 1     | 普段(月曜日から金曜日)、1日当たりどれくらいの時間、テレビやビデオ・DVDを見たり、聞いたりしますか | ①全く見たり、聞いたりしない<br>②1時間より少ない<br>③1時間以上、2時間より少ない<br>④2時間以上、3時間より少ない<br>⑤3時間以上、4時間より少ない | SA   | H28からH29にかけて選択枝に変更あり                             |
| ...  | ...   | ...   | ...  | ...  | ...  |
| 300  | 1     | 友達に会うのは楽しいと思いませんか                                   | ①思う<br>②思わない   | SA   | H15からH16にかけて調査項目名に変更あり<br>H23からH24にかけて調査項目名に変更あり |
| ...  | ...   | ...   | ...  | ...  | ...  |

回答形式を一覧表からコピー＆ペーストする

⑧ 調査項目番号（設問番号）の記入

- 回答形式の右隣の列に、接続対象とする時点分の列を追加し、各時点の調査票における調査項目番号（設問番号）を記入します。このとき、同じ調査項目の同じバージョンとして接続したい年度は同一行に記載しますが、バージョンが異なる場合は、それぞれのバージョンが該当する年度の列のみ調査項目番号（設問番号）を記載します。バージョンが異なる年度や調査項目を実施していない年度の列には、「-」（ハイフン）を記入します。

| 整理番号 | バージョン | 調査項目名<br>(設問文)      | 選択枝  | 回答形式 | 調査項目番号<br>(設問番号) |     |     |     |     | 備考欄                  |
|------|-------|---------------------|--|------|------------------|-----|-----|-----|-----|----------------------|
|      |       |                     |  |      | H19              | ... | R4  | R5  | R6  |                      |
| 1    | 1     | 朝食を毎日食べていますか        | ①している<br>②どちらかといえば、している<br>③あまりしていない<br>④全くしていない | SA   | 1                | ... | 1   | 1   | 1   |                      |
| 2    | 1     | 毎日、同じくらいの時刻に寝ていますか  | ①している<br>②どちらかといえば、している<br>③あまりしていない<br>④全くしていない | SA   | 4                | ... | 2   | 2   | 2   |                      |
| 3    | 1     | 毎日、同じくらいの時刻に起きていますか | ①している<br>②どちらかといえば、している                          | SA   | 5                | ... | 3   | 3   | 3   |                      |
| 4    | 1     | 授業で学んだ生活に特にかしりありますか | ①している<br>②どちらかといえば、している<br>③あまりしていない<br>④全くしていない | SA   | -                | ... | 4   | 4   | 4   | R3からR4にかけて調査項目名に変更あり |
| 5    | 1     | 分からないがあったときも、工夫しますか | ①している<br>②どちらかといえば、している                          | SA   | -                | ... | -   | -   | 5   |                      |
| ...  | ...   | ...                 | ...  | ...  | ...              | ... | ... | ... | ... | ...                  |

調査期間分の列を設け、各時点における調査項目番号を記入する  
実施していない時点やバージョンが異なる時点は「-」を記入する

- これで調査項目マスタの作成は完了です。

✔ Point 調査項目マスタ作成時のポイント

- 調査項目マスタの作成手順④⑥において、同じ調査項目（一覧表作成時に同内容又は概ね似た内容と整理した場合）であっても一部の単語が途中から変更・追加されているなど、差異が生じているものについては、全く同じ調査項目として接続するべきか判断する必要があります。
- 基本的には、調査項目名や選択肢の差異によって**調査項目の趣旨が変わらない場合は接続し、調査項目の趣旨が変わる場合は接続せず、異なるバージョンとして扱います。**
- 同じ調査項目として接続する場合の具体例と、「調査項目マスタ」への反映方法は以下の通りです。

句読点の使い方が異なる場合

➡統一基準とする時点の調査項目における表現・表記を記載します

| 調査時点 1         | 調査時点 2        | マスタへの反映       |
|----------------|---------------|---------------|
| 学級編成の基準は、何人ですか | 学級編成の基準は何人ですか | 学級編成の基準は何人ですか |

漢字・ひらがな・カタカクの表記揺れがある場合

➡統一基準とする時点の調査項目の表現・表記を記載又は正しい表現・表記を記載します

| 調査時点 1  | 調査時点 2   | マスタへの反映  |
|---|--|--|
| …実験などの活動を通して、数量や図形などの性質を <b>見いだす</b> 活動を行いましたか。 | …実験などの活動を通して、数量や図形などの性質を <b>見出す</b> 活動を行いましたか。 | …実験などの活動を通して、数量や図形などの性質を <b>見出す</b> 活動を行いましたか。 |

漢数字と算用数字が混在している場合

➡統一基準とする時点の調査項目の表現・表記を記載又は正しい表現・表記を記載します

| 調査時点 1                     | 調査時点 2                             | マスタへの反映                            |
|----------------------------|------------------------------------|------------------------------------|
| 一学級あたりの生徒数は、次のどれにあてはまりますか。 | <b>1</b> 学級あたりの生徒数は、次のどれにあてはまりますか。 | <b>1</b> 学級あたりの生徒数は、次のどれにあてはまりますか。 |

全角・半角が混在している場合

➡半角を記載します

| 調査時点 1                               | 調査時点 2                              | マスタへの反映                            |
|--------------------------------------|-------------------------------------|------------------------------------|
| <b>I C T</b> を活用した校務の効率化（事務の軽減）に取り組ん | <b>ICT</b> を活用した校務の効率化（事務の軽減）に取り組んで | <b>ICT</b> を活用した校務の効率化（事務の軽減）に取り組組 |

ていますか

いますか。

んでいますか。

### 大文字・小文字が混在している場合

→統一基準とする時点の調査項目の表現・表記を記載又は正しい表現・表記を記載します

| 調査時点 1  | 調査時点 2  | マスタへの反映   |
|---|---|---|
| ①家庭訪問により行った<br>②電話や <b>fax</b> により行った<br>③郵便により行った… | ①家庭訪問により行った<br>②電話や <b>FAX</b> により行った<br>③郵便により行った… | ①家庭訪問により行った<br>②電話や <b>FAX</b> により行った<br>③郵便により行った… |

### 調査項目名（設問文）の例示が異なるが、調査項目の趣旨が変わらない場合

→統一基準とする時点の調査項目の表現・表記を記載します

| 調査時点 1   | 調査時点 2   | マスタへの反映  |
|--|--|--|
| 学習の中で <u>コンピュータなどの ICT 機器</u> を使うのは勉強の役に立つと思いますか | 学習の中で <u>PC・タブレットなどの ICT 機器</u> を使うのは勉強の役に立つと思いますか | 学習の中で <u>PC・タブレットなどの ICT 機器</u> を使うのは勉強の役に立つと思いますか |

### 選択肢の表現が異なるが、回答内容に影響しない場合

→統一基準とする時点の調査項目における表現を記載します

| 調査時点 1  | 調査時点 2  | マスタへの反映   |
|---|---|---|
| (設問文)国語の授業で目的に応じて資料を読み、自分の考えを話したり、書いたりしていますか<br>(選択肢)<br><u>①している</u><br><u>②どちらかといえば、している</u><br><u>③どちらかといえば、していない</u><br><u>④していない</u> | (設問文)国語の授業で目的に応じて資料を読み、自分の考えを話したり、書いたりしていますか<br>(選択肢)<br><u>①当てはまる</u><br><u>②どちらかといえば、当てはまる</u><br><u>③どちらかといえば、当てはまらない</u><br><u>④当てはまらない</u> | (設問文)国語の授業で目的に応じて資料を読み、自分の考えを話したり、書いたりしていますか<br>(選択肢)<br><u>①当てはまる</u><br><u>②どちらかといえば、当てはまる</u><br><u>③どちらかといえば、当てはまらない</u><br><u>④当てはまらない</u> |

### 選択肢の順番が異なるが、内容が同じ場合

→統一基準とする時点の調査項目における順番で記載します

| 調査時点 1 | 調査時点 2 | マスタへの反映 |
|--------|--------|---------|
|--------|--------|---------|

|  |  |  |
|--|--|--|
| <p>(設問文)携帯電話やスマートフォンで通話やメールをしていますか</p> <p>(選択肢)</p> <p><u>①携帯電話を持っていない</u></p> <p><u>②全く、または、ほとんどしていない</u></p> <p><u>③ときどきしている</u></p> <p><u>④ほぼ毎日している</u></p> | <p>(設問文)携帯電話やスマートフォンで通話やメールをしていますか</p> <p>(選択肢)</p> <p><u>①ほぼ毎日している</u></p> <p><u>②ときどきしている</u></p> <p><u>③全く、または、ほとんどしていない</u></p> <p><u>④携帯電話を持っていない</u></p> | <p>(設問文)携帯電話やスマートフォンで通話やメールをしていますか</p> <p>(選択肢)</p> <p><u>①ほぼ毎日している</u></p> <p><u>②ときどきしている</u></p> <p><u>③全く、または、ほとんどしていない</u></p> <p><u>④携帯電話を持っていない</u></p> |
|--|--|--|

**選択肢の範囲が異なるが、いずれかの範囲に合わせて修正できる場合**

➡選択肢の範囲を修正した上で、同じ調査項目として扱います。データ接続の際に選択肢番号の修正が必要であることを記録します。

| 調査時点 1   | 調査時点 2  | マスタへの反映   |
|--|---|---|
| <p>(選択肢)</p> <p><u>①0分</u></p> <p><u>②30分程度</u></p> <p><u>③1時間程度</u></p> <p><u>④2時間程度</u></p> <p><u>⑤3時間程度</u></p> | <p>(選択肢)</p> <p><u>①0分</u></p> <p><u>②15分程度</u></p> <p><u>③30分程度</u></p> <p><u>④1時間程度</u></p> <p><u>⑤2時間程度</u></p> <p><u>⑥3時間程度</u></p> | <p>(選択肢)</p> <p><u>①0分</u></p> <p><u>②15分程度</u></p> <p><u>③30分程度</u></p> <p><u>④1時間程度</u></p> <p><u>⑤2時間程度</u></p> <p><u>⑥3時間程度</u></p> <p>調査時点 1 の回答については、データ接続時に以下の処理を行うことを記録しておきます。</p> <p><u>①0分</u></p> <p><u>②(空欄にする)</u></p> <p><u>③30分程度</u></p> <p><u>④1時間程度</u></p> <p><u>⑤2時間程度</u></p> <p><u>⑥3時間程度</u></p> |

## 3-2. 調査項目ごとに識別 ID を付与する

|    |  |
|----|--|
| 目的 | 調査項目ごとに識別 ID を付与する   |
|    | <p>✔ Point</p> <p>データの参照・集計を円滑に行えるよう、各調査項目に固有の識別 ID を付与する必要があります。調査項目の識別 ID の採番ルールを決め、各調査項目に識別 ID を付与していきましょう。</p> |

### 3-2-1.手順③：調査項目の識別 ID の採番ルールを決定し、各調査項目に付与する

- 採番ルールは対象の種類や識別すべき情報に応じて決定する必要があり、厳密な決まりがあるわけではありません。
- 最も簡単な採番方法は、「調査項目マスタ」に一覧化された調査項目に対して順に識別 ID を割り振っていくことです。ここでは先頭にアルファベットをつける場合を用いて説明します。一般的に識別 ID には以下のような情報を含めます。

S (アルファベット) + 00001, 00002, … (数字)

調査項目の識別文字                      調査項目の識別番号

調査の種類・調査名・調査対象の  
頭文字など

桁数を揃えた連続した数字など

(例：児童生徒「Student」の「S」)

|               |  |
|---------------|--|
| 調査項目の<br>識別文字 | <ul style="list-style-type: none"><li>● 調査の種類や回答対象に対応したアルファベット（1～2 文字）を先頭に付与します。特に一つの調査に複数の調査票がある（調査の種類や回答対象が異なる）場合に有用です。<br/>例) 学校 (<b>S</b>chool) — 生徒 (<b>ST</b>udent)<br/>保護者 (<b>P</b>arent) — 子ども (<b>C</b>hild)</li><li>● 調査票が一つの場合のみは、調査名や調査対象を表すアルファベットを用いることもできます。<br/>例) 児童生徒を対象とした調査であれば「Student」の「S」を、<br/>教員を対象とした調査であれば「Teacher」の「T」を付与</li></ul> |
| 調査項目の         | <ul style="list-style-type: none"><li>● 調査項目ごとに番号を付与します。数字の桁数自体については特に決まり</li></ul>  |

## 識別番号

はありませんが、将来的に新しい調査項目が加わることを考慮し、先頭に 0 を付けて桁数を揃えます。

例) 00001, 00002, …, 01234

- あまり少ない桁数を設定してしまうと、調査項目が増えた際に、識別 ID をすべて付与し直す必要が生じる可能性があります。桁数が途中で変更されることで、Excel や各種統計ソフトで意図した調査項目の並びやデータ構造が崩れてしまい、データ管理や分析の効率が低下することも考えられます。そのため、ある程度余裕をもった桁数（例えば、5 桁や 10 桁など、切りが良く、余裕のある桁数）を設定します。

### ✔ Point 識別 ID の採番に関する一般的なルール

- Excel や各種統計ソフトで文字化けする可能性があるため、漢字やひらがな・カタカナは使用しない。
  - アルファベットの大文字・小文字の併用を控え、どちらかに統一する。
  - アルファベット・数字ともに半角にする。
  - 仮に記号を用いる場合にはピリオド (.) の使用を控え、アンダーバー ( \_ ) を使用する。
  - どのような方針に基づいて採番・付与したのかが分かるよう、一貫したルールを定めて記録しておく。
- 
- 採番ルールを決定したら、調査項目マスタの 1 行目から順に識別 ID を付与していきます。一覧表と対応させるために、整理番号を残しています。

| 識別ID   | 整理番号 | バージョン | 調査項目名<br>(設問文)                  | 選択肢   | 回答形式 | 調査項目番号<br>(設問番号) |     |     |     | 備考欄                    |
|--------|------|-------|---------------------------------|---|------|------------------|-----|-----|-----|------------------------|
|        |      |       |                                 |   |      | H19              | H20 | ... | R6  |                        |
| S00001 | 1    | 1     | 朝食を毎日食べていますか                    | ①している<br>②どちらかといえば、している<br>③あまりしていない<br>④全くしていない            | SA   | 1                | 1   | ... | 1   |                        |
| ...    | ...  | ...   | ...                             | ...   | ...  | ...              | ... | ... | ... |                        |
| S00108 | 108  | 1     | 地域や社会をよくするために何をすべきかを考えることがありますか | ①当てはまる<br>②どちらかといえば、当てはまる<br>③どちらかといえば、当てはまらない<br>④当てはまらない  | SA   | 29               | 30  | ... | -   |                        |
| S00109 | 108  | 2     | 地域や社会をよくするために何かしてみたいと思いますか      | ①当てはまる<br>②どちらかといえば、当てはまる<br>③どちらかといえば、当てはまらない<br>④当てはまらない  | SA   | -                | -   | ... | 25  |                        |
| ...    | ...  | ...   | ...                             | ...   | ...  | ...              | ... | ... | ... |                        |
| S00140 | 137  | 1     | 携帯電話やスマートフォンで通話やメールをしていますか      | ①携帯電話を持っていない<br>②全く、または、ほとんどしていない<br>③ときどきしている<br>④ほぼ毎日している | SA   | 19               | 19  | ... | -   | H19からH20にかけて調査項目名に変更あり |

### Point 政府統計の総合窓口 (e-Stat) における識別 ID

- 政府統計の総合窓口 (e-Stat) では、調査項目の識別 ID として、調査票ごとに「調査項目 ID」を付与しています。

調査票の調査項目一覧

|         |                     |
|---------|---------------------|
| 政府統計コード | 00200521            |
| 調査実施機関  | 総務省                 |
| 政府統計名   | 国勢調査                |
| 調査票 ID  | 00101               |
| 調査票     | 調査票 (大規模調査) (平成12年) |
| 備考      |                     |

調査票のイメージを表示    調査票定義のダウンロード

調査項目

1/2ページ [全28件]

| 調査項目ID    | 調査項目名         | 定義単位項目ID | 定義単位項目名  |
|-----------|---------------|----------|----------|
| 001010001 | 男女の別          | 35770    | 男女の別     |
| 001010002 | 世帯主との続き柄      | 32680    | 世帯主との続き柄 |
| 001010003 | 出生時期          | 31970    | 出生時期     |
| 001010004 | 配偶の関係         | 39981    | 配偶の関係_1  |
| 001010005 | 国籍            | 23620    | 国籍       |
| 001010006 | 現在の住居における居住期間 | 19320    | 居住期間     |

(出所) 政府統計の窓口 (e-Stat) 「調査票の調査項目一覧」

(<https://www.e-stat.go.jp/surveyitems/surveyforms/00101>)

- 「調査票（大規模調査）（平成 12 年）」の調査票 ID を「00101」とし、個人の性別を問う調査項目を「0001」としているため、これらを組み合わせて「001010001」のように調査項目 ID を付与しています。



## 作業成果

- **調査項目の識別 ID の採番ルール**
  - 識別 ID として文字と数字の組み合わせを用いるか、数字のみを用いるかを決定する。
    - 文字を用いる場合**
      - 調査の種類や回答対象に対応したアルファベット（1～2 文字）を決定する。
      - アルファベットの後に続く識別番号の桁数を決定する。
    - 数字のみを用いる場合**
      - 識別番号の桁数を決定する。
  - 調査項目マスタの 1 行目から順に識別 ID を付与する。
- **調査項目の識別 ID が記載された「調査項目マスタ」**

### 3-2-2.手順④：調査項目マスタを用いて変数対応表を作成する

- 調査項目の識別 ID が記載された調査項目マスタを用いて、各時点のデータセットの調査項目（以降、変数）に調査項目の識別 ID を割り当てます。その準備として、**変数対応表を作り、各時点の変数名、調査項目の識別 ID 等の対応を記録**します。
- これにより、ここまでで同じ調査項目であると整理したにもかかわらず、**各時点のデータセット間で変数名が異なる場合でも、調査項目の識別 ID を用いて同じものとして紐づける**ことができます。
- 各時点でデータセットの変数名が異なる場合は、統一した「（新）変数名」を設定します。調査項目マスタの調査項目名（設問文）を使用するか、下図のように、各時点で共通した変数名のパターンがあれば、そちらを使用します。
- 実施年や、後述する接続キー等の回答者属性に関する変数名については、例外的に「year」「pref\_code」等の意味のある文字列（単語の組み合わせ等）から構成される、わかりやすい識別 ID を設定することも考えられます。これらの項目は、接続時や分析時に用いる頻度が高いため、わかりやすく意味のある識別 ID とすることが一案となります。

| 新変数名      | 調査項目の識別 ID | R1_□-データ     | R2_□-データ   | R3_□-データ   | R4_□-データ |
|-----------|------------|--------------|------------|------------|----------|
| 実施年       | year       | R1_実施年       | R2_実施年     | R3_実施年     | 実施年      |
| 都道府県コード   | pref_code  | R1_都道府県コード   | R2_都道府県コード | R3_都道府県コード | 都道府県コード  |
| 都道府県名     | pref_name  | R1_都道府県名     | R2_都道府県名   | R3_都道府県名   | 都道府県名    |
| 市町村コード    | city_code  | R1_所在地市町村コード | R2_市町村コード  | R3_市町村コード  | 市町村コード   |
| 市町村名      | city_name  | R1_所在地市町村名   | R2_市町村名    | R3_市町村名    | 市町村名     |
| 所属人数      | S_00001    | R1_所属人数      | R2_所属人数    | R3_所属人数    | 所属人数     |
| 敷地面積      | S_00002    | R1_敷地面積      | R2_敷地面積    | R3_敷地面積    | 敷地面積     |
| 建物構造      | S_00003    | R1_建物形態      | R2_建物形態    | R3_建物構造    | 建物構造     |
| 第1期スコア    | S_00004    | R1_第1期スコア    | R2_第1期スコア  | R3_第1期スコア  | 第1期スコア   |
| 第2期スコア    | S_00005    | R1_第2期スコア    | R2_第2期スコア  | R3_第2期スコア  | 第2期スコア   |
| 貸上げを実施したか | S_00006    | R1_設問01      | —          | R3_設問02    | 設問02     |
| ...       | ...        | ...          | ...        | ...        | ...      |

回答者属性の変数については、見た際にわかりやすくするため、識別IDを意味のある文字列（単語の組み合わせ等）としてもよい

項目が取得されていない時点は統一のルール（ハイフンなど）で記録する

### 3-2-3.手順⑤：変数対応表を用いてデータセットに調査項目の識別IDを反映する

- 変数対応表の「新変数名」及び「調査項目の識別ID」の情報を用いて、各時点のデータセット内の元の変数名を、調査項目の識別IDに置き換えます。この作業によって、各時点のデータの識別IDが揃い、経年での接続が可能になります。以下の図が置き換えのイメージです。

データセット上で設定されていた変数名（「元変数名」）を新たに設定した「調査項目の識別ID」に置き換える

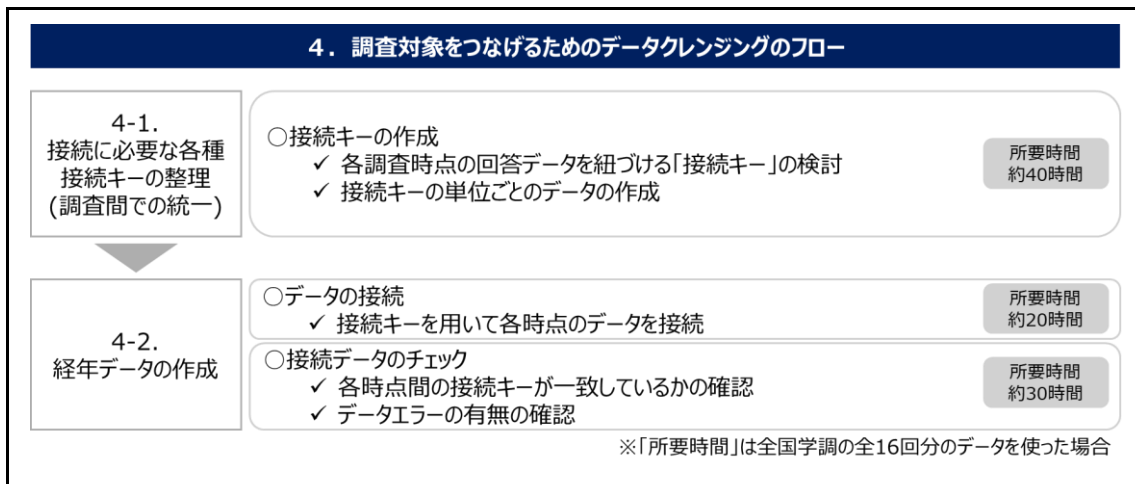
元変数名

| ...  | R1_所属人数 | R1_敷地面積 | R1_施設形態 | R1_第1期スコア | R1_第2期スコア | ...     |         |         |         |     |
|------|---------|---------|---------|-----------|-----------|---------|---------|---------|---------|-----|
| 実施年  | 都道府県コード | 都道府県名   | 市町村コード  | 市町村名      | S_00001   | S_00002 | S_00003 | S_00004 | S_00005 | ... |
| 2019 | 1       | 北海道     | 1〇〇〇    | 〇〇市       | 9.982     | 10.129  | 71.302  | 72.347  | 1.623   | ... |
| 2019 | 1       | 北海道     | 1△△△    | △△市       | 8.200     | 8.796   | 61.954  | 65.183  | 1.391   | ... |
| 2019 | 1       | 北海道     | 1▽▽▽    | ▽▽市       | 7.832     | 8.534   | 58.570  | 62.826  | 1.359   | ... |
| 2019 | 1       | 北海道     | 1□□□    | □□市       | 8.747     | 9.296   | 62.481  | 66.403  | 1.414   | ... |
| 2019 | 1       | 北海道     | 1×××    | ××市       | 8.357     | 8.946   | 59.694  | 63.898  | 1.382   | ... |

- 置き換えが完了した後は、変数対応表と照らし合わせ、正しくデータセットに情報が反映されているか目視で確認します。  
 ➡データセットの確認方法は「4-2-3.手順⑩：接続したデータのチェック（データエラー）を行う」（p.59）へ

## 4. 調査対象をつなげるためのデータクレンジング

第3章までに、複数時点での調査における調査項目をつなげるための準備ができました。続くこの章では、複数時点の調査対象をつなげることを目的としたデータクレンジングについて説明します。



### 4-1. 調査対象の接続に必要な各種接続キーを整理、統一する

#### 目的

データ接続に必要な各種接続キーを整理し、接続対象とする期間の中で統一する

#### ✔ Point

データ接続には、回答データを紐づける接続単位ごとに、「接続キー」が必要になります。時点間で統一した接続キーを用意しましょう。

- パネルデータの場合、同一の回答者を追跡するため、複数時点間で回答者ごとの接続キー（例：回答者 ID やコード）を揃える必要があります。
- 一方、各調査時点で回答者が異なる繰り返しクロスセクションデータの場合、回答者単位ではデータをつなげることができないため、接続が可能な単位（例：学校、事業所、市町村など）にデータをまとめ、複数時点間で、その接続単位ごとに接続キーを揃える必要があります。
- 第2章で触れた調査項目の識別 ID のように、**データ接続を効率化するためには、一意に識別できる ID やコードを接続キーとして使用することが極めて重要**です。一意に識別できる接続キーがなく、接続単位の名称（例：学校名や市町村名）等をそのまま接続キーとして使用する場合、全く同じ名称が複数存在し区別がつかないことや、複数時点で表記ゆれ（例：「〇〇ヶ丘」と「〇〇ヶ丘」など）が発生し接続できないなどの支障が生じます。**データ接続の効率化と信頼性向上のためには、一意の接続キーを設定し、複数時点間で一貫性を保つことが重要**です。

- また、**データの管理や更新、利活用も容易**になります。例えば、学校の名称や所在地が変更された場合でも、接続キーが一貫していれば、過去データとの連携や追跡をスムーズに行うことができます。
- 接続キーの設定手順に入る前にまず、接続単位を特定できる接続キーが各時点のデータセットに含まれているかを確認します。このとき、データセットの状態としては大きく以下の 3 パターンが考えられます。下図では、事業所 ID を接続キーとして例示しています。

|   |  |
|---|--|
| ① 時点間で接続キーが完全に一致している                          | ● 接続キーを用いて各時点のデータ接続が可能<br>➡「4-2.経年データを作成する」(p.56) へ                          |
| ② 部分的に接続キーが設定されている、もしくは接続キーが設定されているが、時点ごとに異なる | ● 時点間で接続キーを揃える必要があり、各時点で接続キーの設定ルールが異なる場合は、同一の接続単位の追跡やルールの統一が必要<br>➡このまま手順⑥⑦へ |
| ③ 接続キーが設定されていない                               | ● 同一の接続単位を追跡できるようにし、一貫した接続キーの設定が必要<br>➡このまま手順⑥⑦へ                             |

部分的に接続キーが揃っていない。接続キーのない時点については、接続キーのある取得時点に合わせて接続キーを設定する必要がある。

接続キーが設定されていない、もしくはすべての時点で異なるキーが設定されている。統一の接続キーを設定する必要がある。

| 年    | 事業所ID   | 事業所名 | 郵便番号     |
|------|---------|------|----------|
| 2021 | EST_001 | AAA  | XXX-XXX1 |
| 2021 | EST_002 | BBB  | XXX-XXX2 |
| 2021 | EST_003 | CCC  | XXX-XXX3 |
| 2021 | EST_004 | DDD  | XXX-XXX4 |
| 2021 | EST_005 | EEE  | XXX-XXX5 |
| 2021 | EST_006 | FFF  | XXX-XXX6 |
| 2022 | EST_001 | AAA  | XXX-XXX1 |
| 2022 | EST_002 | BBB  | XXX-XXX2 |
| 2022 | EST_003 | CCC  | XXX-XXX3 |
| 2022 | EST_004 | DDD  | XXX-XXX4 |
| 2022 | EST_005 | EEE  | XXX-XXX5 |
| 2022 | EST_006 | FFF  | XXX-XXX6 |
| 2023 | EST_001 | AAA  | XXX-XXX1 |
| 2023 | EST_002 | BBB  | XXX-XXX2 |
| 2023 | EST_003 | CCC  | XXX-XXX3 |
| 2023 | EST_004 | DDD  | XXX-XXX4 |
| 2023 | EST_005 | EEE  | XXX-XXX5 |
| 2023 | EST_006 | FFF  | XXX-XXX6 |

| 年    | 事業所ID   | 事業所名 | 郵便番号     |
|------|---------|------|----------|
| 2021 | -       | AAA  | XXX-XXX1 |
| 2021 | -       | BBB  | XXX-XXX2 |
| 2021 | -       | CCC  | XXX-XXX3 |
| 2021 | -       | DDD  | XXX-XXX4 |
| 2021 | -       | FFF  | XXX-XXX6 |
| 2021 | -       | EEE  | XXX-XXX5 |
| 2022 | EST_001 | AAA  | XXX-XXX1 |
| 2022 | EST_002 | BBB  | XXX-XXX2 |
| 2022 | EST_003 | CCC  | XXX-XXX3 |
| 2022 | EST_004 | DDD  | XXX-XXX4 |
| 2022 | EST_005 | EEE  | XXX-XXX5 |
| 2022 | EST_006 | FFF  | XXX-XXX6 |
| 2023 | EST_001 | AAA  | XXX-XXX1 |
| 2023 | EST_002 | BBB  | XXX-XXX2 |
| 2023 | EST_003 | CCC  | XXX-XXX3 |
| 2023 | EST_004 | DDD  | XXX-XXX4 |
| 2023 | EST_005 | EEE  | XXX-XXX5 |
| 2023 | EST_006 | FFF  | XXX-XXX6 |

| 年    | 事業所ID | 事業所名 | 郵便番号     |
|------|-------|------|----------|
| 2021 | -     | CCC  | XXX-XXX3 |
| 2021 | -     | DDD  | XXX-XXX4 |
| 2021 | -     | AAA  | XXX-XXX1 |
| 2021 | -     | BBB  | XXX-XXX2 |
| 2021 | -     | EEE  | XXX-XXX5 |
| 2021 | -     | FFF  | XXX-XXX6 |
| 2022 | -     | CCC  | XXX-XXX3 |
| 2022 | -     | DDD  | XXX-XXX4 |
| 2022 | -     | EEE  | XXX-XXX5 |
| 2022 | -     | AAA  | XXX-XXX1 |
| 2022 | -     | BBB  | XXX-XXX2 |
| 2022 | -     | FFF  | XXX-XXX6 |
| 2023 | -     | AAA  | XXX-XXX1 |
| 2023 | -     | BBB  | XXX-XXX2 |
| 2023 | -     | CCC  | XXX-XXX3 |
| 2023 | -     | DDD  | XXX-XXX4 |
| 2023 | -     | EEE  | XXX-XXX5 |
| 2023 | -     | FFF  | XXX-XXX6 |

#### 4-1-1.手順⑥：一貫した「接続キー」を設定し、データセットに反映する

- データの接続単位が学校や市町村、企業の場合は、「学校コード」、「全国地方公共団体コード<sup>6</sup>」、「法人番号」のように、**行政機関が設定した接続キーを用いる**ことが有効です。これらの接続キーは一貫であることに加え、同様の接続キーが用いられている他の調査と接続したデータ活用ができるメリットもあります。
- 接続単位において行政機関等が設定した接続キーがない場合は、都道府県コードや全国地方公共団体コードを参考にしながら、わかりやすい接続キーを設定することが重要です。
  - ➔接続キーの設定方法や運用について知りたい場合は、「「接続キー」を新たに設定する方法について」(p.50)を参照
- **新しい接続キーと、データの接続単位（例：学校、市町村など）を一覧化した対応表を作成**します。以下は、元の接続キーが部分的に揃っていない場合に、共通した接続キーを設定した対応表のイメージです。
- この対応表をもとに、各時点のデータセットに接続キーの変数を加えます。

##### <接続キーを設定した対応表の例>

共通の接続キーを設定。  
これにより、時点ごとにコードが異なる  
場合にも接続することが可能。

各時点で設定された接続キーが異なっているため、  
統一の接続キーが必要。  
(この例では2018と2021が異なっている)

| 接続キー       | 学校名     | 2018_コード | 2019_コード | 2020_コード | 2021_コード |
|------------|---------|----------|----------|----------|----------|
| N212000123 | A市立a小学校 | 1800101  | 100101   | 100101   | 0100101  |
| N212000112 | A市立b小学校 | 1800102  | 100102   | 100102   | 0100102  |
| N212000124 | A市立c小学校 | 1800103  | 100103   | 100103   | 0100103  |
| N212000145 | A市立d小学校 | 1800104  | 100104   | 100104   | 0100104  |
| N212000198 | A市立e小学校 | 1800105  | 100105   | 100105   | 0100105  |
| N212000125 | B市立f小学校 | 1800107  | 100106   | 100106   | 0100106  |
| N212000143 | B市立g小学校 | 1800108  | 100107   | 100107   | 0100107  |
| N212000130 | B市立h小学校 | 1800106  | 100108   | 100108   | 0100108  |
| N212000152 | B市立i小学校 | 1800109  | 100109   | 100109   | 0100109  |
| N212000178 | B市立j小学校 | 1800110  | 100110   | 100110   | 0100110  |

<sup>6</sup> 全国地方公共団体コード (<https://www.soumu.go.jp/denshijiti/code.html>)

✔ **Point 「接続キー」を新たに設定する方法について**

- 部分的に接続キーが設定されている、もしくは接続キーが設定されているが、時点ごとに異なる場合に、新しく共通した接続キーを設定する必要があります。
- 接続キーの設定や運用方法について詳しく知りたい場合は、以下の資料をご参照ください。

● **デジタル庁「GIF(Government Interoperability Framework)政府相互運用性フレームワーク説明資料」pp.46-47**

([https://www.digital.go.jp/assets/contents/node/basic\\_page/field\\_ref\\_resources/c1bdb4a2-850f-48ec-a938-a8369f820ac1/d47def21/20250325\\_policies\\_development\\_management\\_outline\\_03.pdf](https://www.digital.go.jp/assets/contents/node/basic_page/field_ref_resources/c1bdb4a2-850f-48ec-a938-a8369f820ac1/d47def21/20250325_policies_development_management_outline_03.pdf))

- p.46「ユニークID（識別子）」
- p.47「コード（分類系）」

● **デジタル庁「デジタル社会推進実践ガイドブック DS-463-1 コード（分類体系）導入実践ガイドブック」<sup>7</sup>**

([https://github.com/JDA-DM/GIF/tree/v2.1/460\\_%E5%AE%9F%E8%B7%B5%E3%82%AC%E3%82%A4%E3%83%89%E3%83%96%E3%83%83%E3%82%AF](https://github.com/JDA-DM/GIF/tree/v2.1/460_%E5%AE%9F%E8%B7%B5%E3%82%AC%E3%82%A4%E3%83%89%E3%83%96%E3%83%83%E3%82%AF))

- また、名称等の類似度を計算し、同一であるかどうかを判定することで接続キーを整理する「**名寄せ**」（表記ゆれや違う名称で登録されているデータを、同じ対象としてまとめる作業）の方法もあります。  
➔詳細は、「名寄せのプロセス」（p.52）を参照

**4-1-2.手順⑦：接続単位ごとのデータを作成する**

- パネルデータの場合には、接続キーによって経年での接続が可能です。一方、**各時点で異なる回答者を対象とする繰り返しクロスセクションデータの場合、接続キーの単位（学校、事業所、市町村等）ごとのデータを作成**する必要があります。
- 具体的には、**接続キーをもとに回答者の回答値の平均値（接続キーごとの平均値）を算出**することで、接続単位の回答値とみなすことができます。

<sup>7</sup> デジタル庁が運用するGitHub（外部サイト）の「DS-400 政府相互運用性フレームワーク（GIF）」ページのリンクを示しています。

### <接続単位の回答値の作成例>

市町村ごとに学校データを平均

| 学校コード | 学校名 | 市町村コード | 市町村名 | 変数A  |
|-------|-----|--------|------|------|
| S0001 | AB校 | ABC01  | A市   | 8.94 |
| S0002 | AC校 | ABC01  | A市   | 8.92 |
| S0003 | AD校 | ABC01  | A市   | 8.95 |
| S0004 | BC校 | ABC02  | B市   | 9.12 |
| S0005 | BD校 | ABC02  | B市   | 8.92 |

| 市町村コード | 市町村名 | 変数A  |
|--------|------|------|
| ABC01  | A市   | 8.94 |
| ABC02  | B市   | 9.02 |
| ABC03  | C町   | 9.09 |
| ABC04  | D村   | 9.35 |
| ABC05  | E市   | 9.39 |

- **これで、経年データを接続するための準備はすべて整いました。** 次の章では、接続キーを用いて、どのようにデータを接続するかについて説明します。

## Point 名寄せのプロセス

### ① 名称の正規化

- 各調査時点のデータに含まれる接続単位の名称を、比較しやすい形に統一します。接続単位の名称の中心となる部分を抽出して比較しやすい形にします。以下は学校を例とした場合の例です。
  - ◇ 表記ゆれの統一：「高等学校／高校」等の表記ゆれを揃える
  - ◇ 書式の統一：全角・半角、スペース等の使い方を統一する
  - ◇ 名称の中心部分の抽出：「〇〇市立△△小学校」から「△△小学校」など、比較に使う部分のみを切り出す
- このプロセスは、第 2 章で説明した基本的なデータクレンジングが実施されている場合、簡略化できます。

### ② 接続キーマスタの作成

- 名寄せの基準となる「接続キーマスタ」を作成します。接続キーマスタとは、各接続単位に一意の接続キーと代表的な名称・住所等の情報を紐付けた照合用のマスタです。接続キーマスタには、後述する名称の類似度の計算に用いる情報を記載します。
- 接続キーマスタに用いるデータには、主に以下の選択肢があります。
  1. 最新調査時点のデータを用いる：名称の変更があった場合に新しい情報を適用でき、コストが低い。最新調査時点のデータが誤っている場合や、最新調査時点に含まれないデータがある場合、接続キーマスタが不完全なものになる可能性がある。
  2. 複数の調査時点のデータをクラスタリング（同一のデータとみなされる複数のデータをまとめること）して統合する：全時点を通じて欠落のない接続キーマスタを作成することができるが、非常にコストが高い。特定の時点のデータの質に左右されにくい。
- 以下は接続キーマスタのイメージです。この場合、仮接続キーとして連続した学校 ID を採番しています。⑤のプロセスで仮接続キーをそのまま接続キーとして利用するか判断します。

| 学校ID | 学校名    | 郵便番号     | 住所         |       |
|------|--------|----------|------------|-------|
| 001  | AA 小学校 | 150-XX01 | 東京都渋谷区神宮前  | ○-○-○ |
| 002  | AB 小学校 | 154-XX02 | 東京都世田谷区桜丘  | △-△-△ |
| 003  | AC 小学校 | 162-XX56 | 東京都新宿区若葉   | ●-●-● |
| 004  | AD 小学校 | 135-XX61 | 東京都江東区豊洲   | ×-×-× |
| 005  | AE 小学校 | 105-XX13 | 東京都港区浜松町   | ★-★-★ |
| 006  | AF 小学校 | 144-XX41 | 東京都大田区羽田   | ○-○-○ |
| 007  | AG 小学校 | 179-XX76 | 東京都練馬区土支田  | △-△-△ |
| 008  | AH 小学校 | 133-XX57 | 東京都江戸川区西小岩 | ●-●-● |
| 009  | AI 小学校 | 1XX-XX01 | 東京都千代田区千代田 | ×-×-× |
| 010  | AJ 小学校 | 105-XX11 | 東京都港区芝公園   | ★-★-★ |
| 011  | AK 小学校 | 108-XX74 | 東京都港区高輪    | ○-○-○ |
| 012  | AL 小学校 | 108-XX72 | 東京都港区白金    | △-△-△ |

### ③ 名称の類似度の計算

- 各調査時点のデータと接続キーマスタのデータを照合し、「同一の接続単位である可能性」を数値（類似度スコア）として計算します。このステップの目的は、目視での確認が必要な候補を絞り込みつつ、明らかに同一と判断できるものを自動的に特定することです。
- 名称だけで照合すると、調査時点ごとの表記ゆれの影響を受けやすいため、類似度の計算を実施する際には、時点ごとに表記のゆれが少ない情報を組み合わせることが有効です。
- 表記のゆれが少ないものとしては、郵便番号や住所が挙げられます。郵便番号と住所を類似度の計算に用いる場合には、郵便番号や住所の一致度を評価して組み合わせることで、類似度のスコアを計算します。スコアの計算には複数の方法があります。ここでは一例を紹介します。郵便番号、名称、住所ごとに一致度を計算し、重みづけをして合計します。
  1. 郵便番号：段階的な一致度を計算  
完全一致（全7桁）→ 一致度 1.0（100%）  
部分一致（上3桁のみ）→ 一致度 0.5（50%）  
不一致 → 一致度 0.0（0%）
  2. 名称・住所：編集距離（レーベンシュタイン距離）による一致度を計算  
一方の文字列をもう一方に変換するのに必要な「1文字の挿入・削除・置換」の最小回数（編集距離）をもとに計算します。  
例) 「△△小学校」と「〇〇市立△△小学校」を比較する場合  
「〇〇市立」の4文字を挿入する必要がある → 編集距離 = 4  
長い方の文字列の文字数 = 10文字  
一致度 =  $1 - 4 \div 10 = 0.6$ （60%）
  3. 重みづけをして合計  
重みの配分は固定ではなく、データの性質に応じて調整します。たとえば名称の表記ゆれが特に大きいデータであれば、郵便番号の重みを高くする対応が考えられます。  
例)  
名称：0.6（一致度） × 0.5（重み） = 0.30  
郵便番号：1.0（一致度） × 0.3（重み） = 0.30  
住所：0.8（一致度） × 0.2（重み） = 0.16  
合計スコア：0.76
- スコアの計算結果を評価します。閾値を設定し、算出したスコアに応じて、閾値以上、閾値以下の2グループに振り分けます。閾値以上は同一、閾値以下は目視確認が必要となります。④のステップで説明します。
- 閾値の設定は固定ではなく、データの性質や許容できるエラー率に応じて調整します。閾値を高くするほど自動判定の精度は上がりますが、目視確認が必要なケースも増えます。例えば、閾値を0.85と設定した場合、類似度計算の結果0.76であれば、閾値以下のグループとなります。

#### ④ スコアに応じて調査対象を確認

- 閾値以上のグループについては、同一と判定して、接続キーマスタに共通の接続キーを追加します。また、調査時点をまたいで、各調査時点のデータに共通の仮接続キーを付与します。以下は各調査時点のデータに仮接続キーを付与したイメージです。

| 年度   | 調査時点ID   | 学校名    | 郵便番号     | 住所               | 候補_学校ID | 名称類似度 | 郵便番号類似度 | 住所類似度 | 重み付きスコア | 閾値判定 >=0.85 | 学校ID (仮接続キー) |
|------|----------|--------|----------|------------------|---------|-------|---------|-------|---------|-------------|--------------|
| 2018 | 2018-001 | AA 小学校 | 150-XX01 | 東京都渋谷区神宮前 ○-○-○  | 001     | 1.00  | 1.00    | 1.00  | 1.00    | ○           | 001          |
| 2018 | 2018-002 | AB 小学校 | 154-XX02 | 東京都世田谷区桜丘 △-△-△  | 002     | 1.00  | 1.00    | 1.00  | 1.00    | ○           | 002          |
| 2018 | 2018-003 | AC 小学校 | 162-XX56 | 東京都新宿区若葉 ●-●-●   | 003     | 1.00  | 1.00    | 1.00  | 1.00    | ○           | 003          |
| 2018 | 2018-004 | AD 小学校 | 135-XX61 | 東京都江東区豊洲 x-x-x   | 004     | 1.00  | 1.00    | 1.00  | 1.00    | ○           | 004          |
| 2018 | 2018-005 | AE 小学校 | 105-XX13 | 東京都港区浜松町 ★-★-★   | 005     | 1.00  | 1.00    | 1.00  | 1.00    | ○           | 005          |
| 2018 | 2018-006 | AF 小学校 | 144-XX41 | 東京都大田区羽田 ○-○-○   | 006     | 1.00  | 1.00    | 1.00  | 1.00    | ○           | 006          |
| 2018 | 2018-007 | AG 小学校 | 179-XX76 | 東京都練馬区土支田 △-△-△  | 007     | 0.83  | 1.00    | 1.00  | 0.92    | ○           | 007          |
| 2018 | 2018-008 | AH 小学校 | 133-XX57 | 東京都江戸川区西小岩 ●-●-● | 008     | 1.00  | 1.00    | 1.00  | 1.00    | ○           | 008          |
| 2018 | 2018-009 | AI 小学校 | 1XX-XX01 | 東京都千代田区千代田 x-x-x | 009     | 1.00  | 1.00    | 1.00  | 1.00    | ○           | 009          |
| 2018 | 2018-010 | AJ 小学校 | 105-XX11 | 東京都港区芝公園 ★-★-★   | 010     | 1.00  | 1.00    | 1.00  | 1.00    | ○           | 010          |

- 閾値以下のグループについては、「名称の一致度は高いが住所にゆれ・誤りがある」「住所の一致度は高いが名称の一致度が低い」「新規校」等の場合が考えられるため、目視での確認が必要となります。
- 目視での確認が必要となった対象については、目視で名称を確認するとともに、必要に応じて Web 上の情報等を用いることで、同一の対象と判断できるかどうかを確認します。
- 確認の結果、同一と判定できる場合、接続キーマスタに共通の仮接続キーを追加し、共通の各調査時点の該当データに仮接続キーを付与します。以下は閾値以下のグループを目視で確認するイメージです。

| 区分    | 年度   | 調査時点ID   | 学校名   | 郵便番号     | 住所             | 候補1_学校名 | 候補1_スコア | 候補2_学校名 | 候補2_スコア | 備考      | 判定 |
|-------|------|----------|-------|----------|----------------|---------|---------|---------|---------|---------|----|
| ★グレー  | 2021 | 2021-X01 | BB小学校 | 136-YY73 | 東京都江東区北砂 x-x-x | BB小学校   | 0.83    | BA小学校   | 0.72    | 住所ゆれ    |    |
| ★低スコア | 2019 | 2019-X03 | BC小学校 | 135-YY64 | 東京都江東区青海 ★-★-★ | BC小学校   | 0.78    | BX小学校   | 0.69    | 新設校の可能性 |    |
| ★低スコア | 2020 | 2020-X02 | BD小学校 | 190-YY12 | 東京都立川市曙町 ○-○-○ | BD小学校   | 0.74    | CD小学校   | 0.71    | 要確認     |    |
| ★グレー  | 2018 | 2018-X04 | BE小学校 | 135-YY62 | 東京都江東区東雲 △-△-△ | BE小学校   | 0.84    | CE小学校   | 0.78    | 表記ゆれ大   |    |

#### ⑤ 抽出された候補の一覧化

- 上記のプロセスを通じて各時点の調査対象について同一のものを整理することができました。接続した結果を以下のように時点ごとに一覧化します。接続キーについて、行政機関が設定した接続キーが存在する場合（p.49 参照）、仮接続キーの代わりに行政機関が設定した接続キーを接続キーとして採用し、仮接続キーは廃止します。
- 廃校・廃業等で行政機関が設定した接続キーが存在しないデータについては、仮接続キーを接続キーとして採用するか、一定のルールに基づいて新たに接続キーを付与します。
- 一覧化した後は、誤った対象が同一とみなされていないかを複数人で確認し、必要に応じて修正します。

<名寄せ後の接続キーを設定した対応表の例>

共通の接続キーを設定。  
これにより、時点ごとに名称が異なる場合にも接続することが可能。

名寄せによって同一の対象を追跡した場合。  
各時点ごとに異なる名称で回答している場合があるため、統一の接続キーが必要。

| 接続キー       | 学校名     | 2018_学校名 | 2019_学校名 | 2020_学校名 | 2021_学校名 |
|------------|---------|----------|----------|----------|----------|
| N212000123 | A市立a小学校 | a小学校     | a小学校     | a小学校     | A市立a小学校  |
| N212000112 | A市立b小学校 | b小学校     | b小学校     | b小学校     | A市立b小学校  |
| N212000124 | A市立c小学校 | A市立c小学校  | A市立c小学校  | A市立c小学校  | A市立c小学校  |
| N212000145 | A市立d小学校 | A市立d小学校  | A市立d小学校  | A市立d小学校  | A市立d小学校  |
| N212000198 | A市立e小学校 | A市立e小学校  | A市立e小学校  | A市立e小学校  | A市立e小学校  |
| N212000125 | B市立f小学校 | f小       | B市立f小学校  | B市立f小    | B市立f小学校  |
| N212000143 | B市立g小学校 | B市立g小学校  | B市立g小学校  | B市立g小学校  | B市立g小学校  |
| N212000130 | B市立h小学校 | B市立h小学校  | B市立h小学校  | B市立h小学校  | B市立h小学校  |
| N212000152 | B市立i小学校 | B市立i小学校  | B市立i小学校  |          | B市立i小学校  |
| N212000178 | B市立j小学校 | B市立j小学校  | B市立j小学校  | j小学校     | j小学校     |

## 4-2. 経年データを作成する

### データを接続して経年データを作成する

#### 目的

#### Point

接続キーを用いて、調査対象（接続単位）をつなげます。正しく接続できたかを確認するために、データの数え上げや記述統計量からエラーがあるかをチェックし、エラーがあれば修正を行います。これによって経年データが完成します。

### 4-2-1. 手順⑧：接続キーを用いてデータを接続する

- まずは各時点のデータを共通する接続キー（例：学校コード）を用いて縦に接続し、**ロング型（縦に広がったデータ型）のデータセット**を作成します。接続キーごとに、調査年順に並ぶようなデータになっています。ここでは、後の分析を見据えて、統計ソフト等でデータを処理しやすいように、ワイド型のデータセット（横に広がったデータ型）ではなく、ロング型のデータセットを作成することが重要です。

接続キーごとに、調査年順に並べる

#### ■各時点のデータセット

| 学校コード | 学校名  | 国公私 | 学校種 | 調査年  | 国語 正答率 |
|-------|------|-----|-----|------|--------|
| S0001 | A中学校 | 国立  | 中学校 | 2015 | 62.4   |
| S0002 | B中学校 | 公立  | 中学校 | 2015 | 58.1   |
| S0003 | C中学校 | 公立  | 中学校 | 2015 | 65.0   |
| S0004 | D中学校 | 公立  | 中学校 | 2015 | 59.3   |

| 学校コード | 学校名  | 国公私 | 学校種 | 調査年  | 国語 正答率 |
|-------|------|-----|-----|------|--------|
| S0001 | A中学校 | 国立  | 中学校 | 2016 | 61.2   |
| S0002 | B中学校 | 公立  | 中学校 | 2016 | 57.9   |
| S0003 | C中学校 | 公立  | 中学校 | 2016 | 64.7   |
| S0004 | D中学校 | 公立  | 中学校 | 2016 | 60.4   |

| 学校コード | 学校名  | 国公私 | 学校種 | 調査年  | 国語 正答率 |
|-------|------|-----|-----|------|--------|
| S0001 | A中学校 | 国立  | 中学校 | 2017 | 63.1   |
| S0002 | B中学校 | 公立  | 中学校 | 2017 | 59.8   |
| S0003 | C中学校 | 公立  | 中学校 | 2017 | 63.8   |
| S0004 | D中学校 | 公立  | 中学校 | 2017 | 57.9   |

#### ■ロング型：調査対象ごとに、データが縦に並んでいる

| 学校コード | 学校名  | 国公私 | 学校種 | 調査年  | 国語 正答率 |
|-------|------|-----|-----|------|--------|
| S0001 | A中学校 | 国立  | 中学校 | 2015 | 62.4   |
| S0001 | A中学校 | 国立  | 中学校 | 2016 | 61.2   |
| S0001 | A中学校 | 国立  | 中学校 | 2017 | 63.1   |
| S0001 | A中学校 | 国立  | 中学校 | 2018 | 63.3   |
| S0001 | A中学校 | 国立  | 中学校 | 2019 | 64.2   |
| ⋮     | ⋮    | ⋮   | ⋮   | ⋮    | ⋮      |
| S0002 | B中学校 | 公立  | 中学校 | 2015 | 58.1   |
| S0002 | B中学校 | 公立  | 中学校 | 2016 | 57.9   |
| S0002 | B中学校 | 公立  | 中学校 | 2017 | 59.8   |
| S0002 | B中学校 | 公立  | 中学校 | 2018 | 59.3   |
| S0002 | B中学校 | 公立  | 中学校 | 2019 | 60.1   |
| ⋮     | ⋮    | ⋮   | ⋮   | ⋮    | ⋮      |
| S0003 | C中学校 | 公立  | 中学校 | 2015 | 65.0   |
| S0003 | C中学校 | 公立  | 中学校 | 2016 | 64.7   |
| S0003 | C中学校 | 公立  | 中学校 | 2017 | 63.8   |
| S0003 | C中学校 | 公立  | 中学校 | 2018 | 65.8   |
| S0003 | C中学校 | 公立  | 中学校 | 2019 | 65.1   |
| ⋮     | ⋮    | ⋮   | ⋮   | ⋮    | ⋮      |
| S0004 | D中学校 | 公立  | 中学校 | 2015 | 59.3   |
| S0004 | D中学校 | 公立  | 中学校 | 2016 | 60.4   |
| S0004 | D中学校 | 公立  | 中学校 | 2017 | 57.9   |
| S0004 | D中学校 | 公立  | 中学校 | 2018 | 56.6   |
| S0004 | D中学校 | 公立  | 中学校 | 2019 | 57.9   |
| ⋮     | ⋮    | ⋮   | ⋮   | ⋮    | ⋮      |
| S0005 | E中学校 | 私立  | 中学校 | 2015 | 55.8   |
| ⋮     | ⋮    | ⋮   | ⋮   | ⋮    | ⋮      |

D中学校のデータが並んでいる。  
学校名や属性情報もすべての  
行に記載される。

- ロング型データのデータセットを作成する際には、Excelのシートの形で格納された各時点データを、コピー&ペーストして一つのシートに接続することで、ロング型データセットにすることができます。なお、各時点のデータ量が多い場合や、時点ごとに異なる列を有している場合は、上述の方法では難しいため、統計ソフトのデータを縦に接続する関数などを用いることで、ロング型データセットを作成しま

す。

### ✔ Point ワイド型のデータセット

- ワイド型のデータセットは、調査対象ごとに項目の情報が並ぶなど、一覧性が高く、直観的にわかりやすいのが特徴ですが、データ加工・分析の観点では扱いが難しいです。
- 接続対象とする時点が多ければ、列も多くなるためデータ構造が膨張しやすいことや、統計ソフトでは列ごとの操作を前提としている一方で、ワイド型では変数が時点ごとに列が分かれるため、計算やデータの加工が難しくなることが挙げられます。

#### ■ ワイド型：調査対象ごとに、データが横に並んでいる

| 学校コード | 学校名  | 国公私 | 学校種 | 国語正答率_2015 | 国語正答率_2016 | 国語正答率_2017 | 国語正答率_2018 | 国語正答率_2019 | ... |
|-------|------|-----|-----|------------|------------|------------|------------|------------|-----|
| S0001 | A中学校 | 国立  | 中学校 | 62.4       | 61.2       | 63.1       | 63.3       | 64.2       | ... |
| S0002 | B中学校 | 公立  | 中学校 | 58.1       | 57.9       | 59.8       | 59.3       | 60.1       | ... |
| S0003 | C中学校 | 公立  | 中学校 | 65.0       | 64.7       | 63.8       | 65.8       | 65.1       | ... |
| S0004 | D中学校 | 公立  | 中学校 | 59.3       | 60.4       | 57.9       | 56.6       | 57.9       | ... |
| S0005 | E中学校 | 私立  | 中学校 | 55.8       | 55.3       | 58.2       | 57.4       | 56.0       | ... |

- データ接続の後は、**データが正しく紐づいているか確認する（接続キーのチェック）**とともに、**データにエラーが含まれていないか確認する（データエラーのチェック）** 必要があります。次頁以降で説明します。

### 📌 作業成果

- **ロング型の経年データ**

#### 4-2-2.手順⑨：接続したデータのチェック（接続キー）を行う

- データが正しく紐づいているか確認するために、**時点間で接続キーが一致しており一意のもの**になっているかを確認します。
- 4-1 で接続キーを実際のデータに反映した際に、何らかのエラーによって正しく紐づかない場合や、接続キーの誤りが見落とされる場合があります。そこで、実際のデータをもとに確認し、正しく紐づいていない場合には、接続キーを修正します。
- 接続キーのチェックを行う際には、複数の項目を用いてチェックします。例えば、「市町村コード」と「市町村名」を使いチェックを行います。そうすることで、市町村名に誤りがあり異なる市町村コードが割り振られている場合や、市町村名は正しいが、誤った市町村コードが割り振られている場合を発見できます。
- 一般的にデータの紐づけをチェックする方法としては、例えば 5 年分のデータを接続する場合、接続キーごとに 5 年分のデータが含まれているかどうかを確認します。確認には、データ内の項目が同じものを数え上げる関数などを用います。
- ここで 5 年分のデータが揃っていない場合は、接続キーに誤りがある可能性があるため、接続キーを用いて検索を行い、目視で誤りが含まれていないか確認を行います。表記ゆれや表記が統一されていない場合も含め、誤りがあった際には、接続キーごとに一意となるように修正を行い、正しく紐づけられるようにします。

| 接続キー  | 市町村名             | count |
|-------|------------------|-------|
| ABC01 | A市               | 5     |
| ABC02 | B市               | 4     |
| ABC21 | B市               | 1     |
| ABC03 | C町               | 3     |
| ABC03 | CC郡C町            | 2     |
| ABC04 | DD村              | 4     |
| ABC04 | D <del>D</del> 村 | 1     |
| ABC05 | E町               | 5     |
| ABC06 | F町               | 5     |
| ...   | ...              | ...   |

市町村コードに誤りがある場合

このデータの場合、出現数が4個以下のデータを確認する

表記ゆれがある場合

表記が統一されていない場合

- 修正を行った場合は、以下のように修正履歴を記録し、管理します。

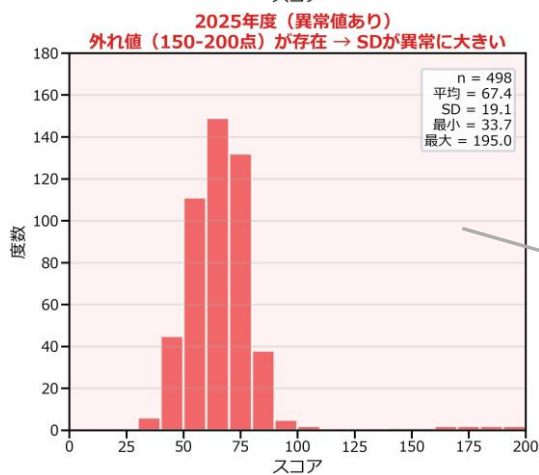
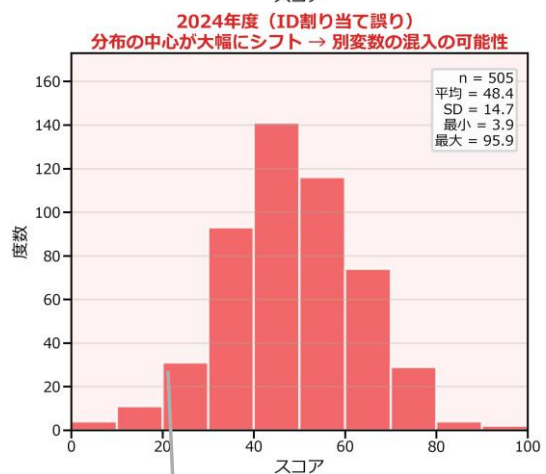
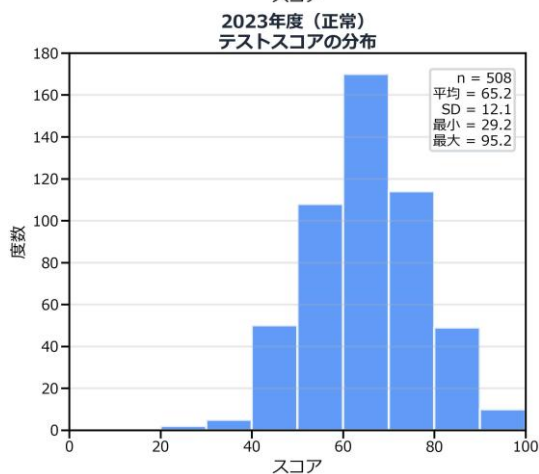
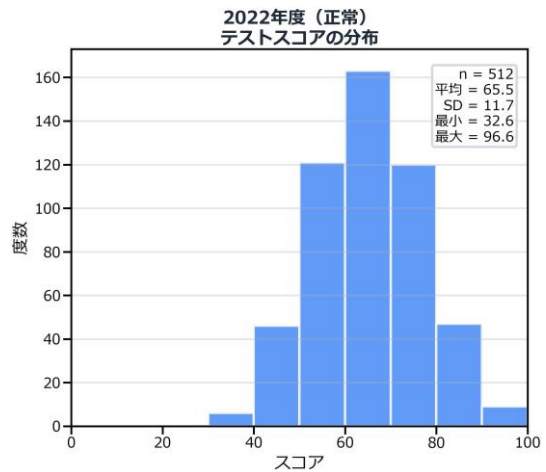
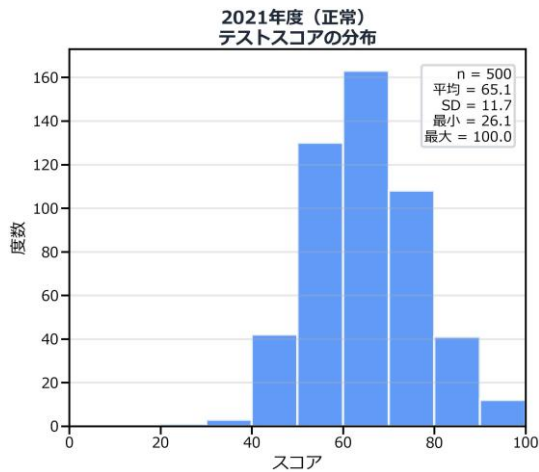
| 実施年       | 接続キー  | 市町村名             | 接続キー_修正 | 市町村_修正 | 備考          |
|-----------|-------|------------------|---------|--------|-------------|
| 2024      | ABC21 | B市               | ABC02   | B市     |             |
| 2023、2024 | ABC03 | CC郡C町            | ABC03   | C町     |             |
| 2021年以外   | ABC04 | D <del>D</del> 村 | ABC04   | DD村    | 2021年の表記に統一 |

### 4-2-3.手順⑩：接続したデータのチェック（データエラー）を行う

- 一般的にデータのエラーをチェックする方法として、**記述統計量（データの基本的な性質を表す指標、平均値や標準偏差、四分位数などのこと）を算出**することや**ヒストグラムを作成**すること等が挙げられます。例えば、5年分のデータを接続した場合を考えます。
- 各時点に変数の記述統計量を算出し、変数がすべての時点で揃っているか確認します。もしデータが欠損している場合、その年度の数値は算出されないため、欠損を見つけることができます。
- 記述統計量やヒストグラムの作成による確認に関しては、時点間で変数の分布が大幅に異なっている場合、調査項目の識別IDの割り当ての誤りや、データに異常値や外れ値が含まれている可能性があります。
- データ内に異常値や外れ値が含まれている場合は、分析の結果が歪んでしまう可能性があるため、一定の基準で欠損値指定を行うこと等を検討する必要があります。

| 実施年                                  | 識別ID        | サンプル数 | 平均    | 標準偏差 | 最小値   | 第1四分位 | 中央値   | 第3四分位  | 最大値    |
|--------------------------------------|-------------|-------|-------|------|-------|-------|-------|--------|--------|
| 2020                                 |             | 19474 | 8.94  | 1.30 | 0.00  | 8.19  | 8.97  | 9.72   | 14.00  |
| 2021                                 |             | 19141 | 9.02  | 1.14 | 0.00  | 8.37  | 9.02  | 9.68   | 14.00  |
| 取得時点のデータが欠損しており、変数がデータに含まれていない可能性がある |             | 18971 | 9.09  | 1.16 | 0.00  | 8.45  | 9.13  | 9.78   | 14.00  |
|                                      |             | 18841 | 9.35  | 1.05 | 0.00  | 8.78  | 9.38  | 9.98   | 14.00  |
|                                      |             | 18700 | 9.39  | 1.13 | 0.00  | 8.78  | 9.42  | 10.06  | 14.00  |
| 2020                                 |             | 19472 | 9.26  | 1.09 | 1.00  | 8.67  | 9.28  | 9.87   | 14.00  |
| 2021                                 |             | 19140 | 11.15 | 1.23 | 0.00  | 10.46 | 11.15 | 11.87  | 16.00  |
| 2022                                 | A110601_000 | NA    | NA    | NA   | NA    | NA    | NA    | NA     | NA     |
| サンプル数が他の取得時点と異なり、データの変更が生じている可能性がある  |             | 18838 | 9.92  | 1.36 | 0.00  | 9.14  | 9.91  | 10.70  | 16.00  |
|                                      |             | 18698 | 10.02 | 1.41 | 0.00  | 9.23  | 10.00 | 10.84  | 16.00  |
|                                      |             | 19474 | 63.83 | 9.32 | 0.00  | 58.53 | 64.09 | 69.46  | 100.00 |
| 2022                                 | A120501_000 | 7850  | 64.40 | 8.12 | 0.00  | 59.77 | 64.42 | 69.14  | 100.00 |
| 2023                                 |             | 18971 | 64.91 | 8.27 | 0.00  | 60.34 | 65.23 | 69.83  | 100.00 |
| 2024                                 |             | 18841 | 66.76 | 7.49 | 0.00  | 62.70 | 66.99 | 71.31  | 100.00 |
| 2024                                 |             | 18700 | 67.05 | 8.04 | 0.00  | 62.72 | 67.30 | 71.83  | 100.00 |
| 2020                                 | A120601_000 | 19473 | 1.51  | 0.30 | 0.00  | 1.34  | 1.51  | 1.68   | 3.00   |
| 2021                                 |             | 19140 | 69.69 | 7.68 | 0.00  | 65.37 | 69.70 | 74.10  | 100.00 |
| 2022                                 |             | 18971 | 62.40 | 8.04 | 0.00  | 57.84 | 62.30 | 67.77  | 100.00 |
| 2023                                 |             | 18838 | 62.00 | 8.52 | 0.00  | 57.13 | 62.00 | 67.77  | 100.00 |
| 2024                                 | 18698       | 62.59 | 8.79  | 0.00 | 57.68 | 62.59 | 67.77 | 100.00 |        |

- 記述統計量の算出には、統計ソフトで記述統計量を算出する関数を用います。データの分布を確認するため、サンプル数、平均値、標準偏差（もしくは分散）、中央値、最小値、最大値などを確認するとよいでしょう。平均値など一つの指標のみでは、分布の歪みや特徴を見落としてしまう可能性があるため、ヒストグラムを作成することや、カテゴリーによって集計する等、複数の指標で確認することが望ましいです。
- 以下はヒストグラムの作成イメージです。時点ごとに変数の分布を可視化することができます。



2024年度データは分布が全く異なっているため、別変数になっている可能性がある

2025年度データは外れ値が含まれるため、標準偏差が大きくなっている

- ここまでのプロセスで、基本的な経年データの作成は完了しました。

**作業成果**

- 接続キーの修正履歴
- 基礎統計量、ヒストグラム
- チェックが終わったロング型の経年データ

## 5. 経年データを用いた分析

- ここでは、経年データを用いた分析手法を紹介するとともに、実証的共同研究における「全国学力・学習状況調査」の経年データを用いた分析事例を紹介します。
- 経年データを用いた分析手法として、以下のものが代表的な例として挙げられます。このうち、グラフによる可視化、クロス集計は Excel での分析も可能ですが、好事例抽出、回帰分析、固定効果モデルなどの分析手法については、R や SPSS 等の統計ソフトを用いることが望ましいです。
- 分析に際しては、「1-6.本手引きで想定するツール」で触れたように、Excel はデータ量が多い（数万行以上）場合の分析に適しておらず、処理速度の速い統計ソフトを用いることを推奨します。

|          | 手法        |   | 統計ソフト (R、SPSS等)      |                 |
|----------|-----------|---|----------------------|-----------------|
|          | 手法        | 分析方法  | Excel                | 統計ソフト (R、SPSS等) |
| 比較的簡便な手法 | グラフによる可視化 | 経年での変化や推移を可視化する。                                      | ○                    | ○               |
|          | クロス集計     | クロス集計を用いて、施策の効果を測定する。                                 | ○                    | ○               |
| 比較的厳密な手法 | 好事例抽出     | 経年での推移や伸びが見られるサンプルを抽出する。                              | △<br>時間はかかるが実施可能     | ○               |
|          | 回帰分析      | 最小二乗法を用いて、施策の効果を測定する。                                 | △<br>応用性が低いが実施可能     | ○               |
|          | 固定効果モデル   | パネルデータの分析で使用される手法であり、個体間の違いによる影響を制御し、変数間の関係をより正確に捉える。 | ×<br>関数が含まれておらず実施不可能 | ○               |

- 以下では、「全国学力・学習状況調査」の経年データを用いた分析事例を紹介します。

### 「全国学力・学習状況調査」の分析事例を読み解く際の留意点

- 「全国学力・学習状況調査」の結果データはパネルデータではないため、厳密な因果推論を行うことはできませんが、サンプルサイズが大きく、かつ長期間のデータが取得されていることから、トレンドの把握が可能となっています。
- また、実証的共同研究で分析に用いた年度の学力データは経年比較のために教科調査の難易度を厳密に調整する設計になっていないため、正答率そのものによる経年比較は適当ではありませんでした。

- 以上を踏まえ、実証的共同研究では、正答率の「標準化スコア<sup>8</sup>」を代替的に用いることで、集団内での相対的な位置を比較しました。  
 →詳しい分析結果については、「全国学力・学習状況調査結果の長期的分析等に係る調査研究の請負 報告書 概要版」(pp.21-57)を参照

### 分析事例 1 : 特徴的な成果を示す学校・市町村の抽出

|             |   |
|-------------|---|
| 分析目的        | ポジティブデビアンズ（同じ悪条件下で突出した成果を示すもの）に着目し、SES（社会経済的地位）が低い環境にありながら、高い（＝他の市町村・学校と比較して相対的な位置が高い）成果をあげている学校・市町村を把握する。    |
| 分析結果の活用イメージ | 抽出した学校や市町村（教育委員会）に追加的に質的な調査（ヒアリング等）を行うことを想定。データ上だけでは把握できない、特徴的な取組の背景や具体的な実践事例を詳細に捉え、各校・地域への指導・助言に活かすことを期待     |
| 分析単位        | 学校単位及び市町村単位   |
| 対象期間        | 平成 31 年～令和 6 年  |
| 分析内容        | SES が相対的に低い市町村を地域規模別（大規模/中規模/小規模）に分け、国語・算数（数学）の正答率の標準化スコアが継続的に高い又は伸びを示す市町村を抽出した。抽出された市町村から、同様の特徴が見られる学校を抽出した。 |

### 分析事例 2 : 固定効果モデルを用いた分析

|             |  |
|-------------|--|
| 分析目的        | 学力の向上に効果的な施策・取組を把握する。  |
| 分析結果の活用イメージ | 分析結果から、学力の向上等に寄与する施策や取組を把握し、政策の評価や今後の政策検討、各学校・地域への指導・助言に活かすことを期待                                   |
| 分析単位        | 学校単位   |
| 対象期間        | 平成 31 年～令和 6 年   |
| 分析内容        | 「学校運営の改善や教職員の資質向上につながる取組が見られる学校では、学力の向上（＝集団内での相対的な位置の向上）が見られるのではないか」という分析仮説を設定し、固定効果モデルを用いて分析を行った。 |

- このように、経年データを用いた分析を行うことで、これまで一時点の調査結果や一部のデータだけでは把握できなかった変化や傾向、関係性を、より長期的かつ多面的に明らかにすることが可能になります。このことは、政策効果の内容や程度・影響範囲の把握を行い、政策の振り返りや改善の検討に役立ちます。

<sup>8</sup> 平均 0、標準偏差 1 の尺度に変換した値を指す。

## ■ データクレンジング及び経年データ作成のチェックリスト

| No. | 工程            | 確認事項   |
|-----|---------------|--|
| 1   | 基本的なデータクレンジング | <p><b>データ構造の把握 (⇒2-1 (pp.15-16) )</b></p> <ul style="list-style-type: none"> <li><input type="checkbox"/> 一時点データの構造 (回答取得時点・期間情報/識別情報/属性情報/回答情報) を把握した</li> </ul> <p><b>基本的なデータクレンジングの実施 (⇒2-2 (pp.17-20) )</b></p> <ul style="list-style-type: none"> <li><input type="checkbox"/> 数値の全角・半角表記が統一されていることを確認した</li> <li><input type="checkbox"/> 同じ内容の表記ゆれが統一されていることを確認した</li> <li><input type="checkbox"/> 漢数字と算用数字の表記が統一されていることを確認した</li> <li><input type="checkbox"/> アルファベットの大文字・小文字表記が統一されていることを確認した</li> <li><input type="checkbox"/> 誤字・脱字が修正されていることを確認した</li> <li><input type="checkbox"/> 重複データが削除されていることを確認した</li> <li><input type="checkbox"/> 欠損値 (無回答や非該当) の表記が統一されていることを確認した</li> <li><input type="checkbox"/> 無回答と非該当が区別されて表記されていることを確認した</li> <li><input type="checkbox"/> データが Excel や CSV など機械による判読が可能な形式で保存されていることを確認した</li> <li><input type="checkbox"/> 1セルに1つのデータが記録されていることを確認した</li> <li><input type="checkbox"/> 数値データに文字列や不要なスペースが含まれていないことを確認した</li> <li><input type="checkbox"/> 日付などのデータが機械判読可能な形式で入力されていることを確認した</li> <li><input type="checkbox"/> 数値データに文字列が含まれていないことを確認した</li> <li><input type="checkbox"/> 改行や不要なスペースがデータ内に含まれていないことを確認した</li> <li><input type="checkbox"/> 「セルの結合」機能が使用されていないことを確認した</li> </ul> <p><b>活用目的・作業方針の確認 (⇒2-3 (pp.20-21) )</b></p> <ul style="list-style-type: none"> <li><input type="checkbox"/> 経年データの活用目的を明確にした</li> <li><input type="checkbox"/> 接続単位 (個人、学校、市町村等) を決定した</li> <li><input type="checkbox"/> 接続対象期間を決定した</li> <li><input type="checkbox"/> 接続するデータの範囲を決定した</li> </ul> <p><b>接続に向けたデータ概要の確認 (⇒2-4 (pp.21-24) )</b></p> <ul style="list-style-type: none"> <li><input type="checkbox"/> 複数時点分の調査データが揃っているか確認した</li> <li><input type="checkbox"/> 各調査時点において調査対象が同じであることを確認した (同一個人等への追跡調査である必要はない)</li> <li><input type="checkbox"/> 各調査時点において調査内容が概ね同じであることを確認した</li> <li><input type="checkbox"/> 調査の実施頻度や実施時期に大きなばらつきがないか確認した</li> </ul> |

|   |                       |   |
|---|-----------------------|---|
| 2 | 調査項目をつなげるためのデータクレンジング | <p><b>調査項目の一覧化 (⇒3-1 (pp.25-41))</b></p> <ul style="list-style-type: none"> <li><input type="checkbox"/> 各時点の調査項目 (設問文・回答形式・選択肢) の一覧表を作成した</li> <li><input type="checkbox"/> 「調査項目マスタ」を作成し、各調査項目の変更を整理した</li> </ul> <p><b>調査項目の識別 ID の付与 (⇒3-2 (pp.42-46))</b></p> <ul style="list-style-type: none"> <li><input type="checkbox"/> 調査項目の識別 ID の採番ルールを決め、各調査項目に ID を付与した</li> <li><input type="checkbox"/> 調査項目の識別 ID を各時点のデータセットに反映した</li> </ul>  |
| 3 | 調査対象をつなげるためのデータクレンジング | <p><b>接続キーの設定 (⇒4-1 (pp.47-55))</b></p> <ul style="list-style-type: none"> <li><input type="checkbox"/> 接続単位ごとに必要な「接続キー」を整理した</li> <li><input type="checkbox"/> 既存の接続キー (学校コード、市町村コード等) が利用可能か確認した</li> <li><input type="checkbox"/> 既存の接続キーがない場合、新たに接続キーを作成した</li> <li><input type="checkbox"/> 接続キーの対応表を作成し、複数時点間で一貫性を確保した</li> <li><input type="checkbox"/> 接続キーに抜け漏れや重複がないか確認した</li> <li><input type="checkbox"/> 必要に応じて、データの集計 (学校単位から市町村単位への変更等) を行った</li> </ul> <p><b>経年データの作成 (⇒4-2-1 (pp.56-57))</b></p> <ul style="list-style-type: none"> <li><input type="checkbox"/> 接続キーを用いて各時点のデータを縦に接続 (ロング型データ化) した</li> </ul> <p><b>エラーのチェック (⇒4-2-2、4-2-3 (pp.58-60))</b></p> <ul style="list-style-type: none"> <li><input type="checkbox"/> 接続後のデータにエラー (欠損値・異常値・重複等) がないか確認した</li> <li><input type="checkbox"/> 記述統計量 (平均・標準偏差・最小値・最大値等) でデータの分布を確認した</li> <li><input type="checkbox"/> 接続後のデータで、接続キーが一意に紐づいているか確認した</li> </ul> |
| 4 | 記録・再現性の確保             | <p><b>記録・引継ぎ (⇒各手順)</b></p> <ul style="list-style-type: none"> <li><input type="checkbox"/> データクレンジングやデータ接続の作業履歴を記録した</li> <li><input type="checkbox"/> 設問文・選択肢・接続キー等の変更履歴を記録した</li> <li><input type="checkbox"/> 作業手順やルールを文書化し、関係者間で共有した</li> </ul>  |

## ■ 参考資料

---

- 九州工業大学数理・DS・AI 教育推進室「第 5 回 前処理・クレンジング処理」  
(<https://kyutech.repo.nii.ac.jp/records/2001414>)
- 経済産業省 大臣官房調査統計グループ 統計企画室「令和 4 年度統計調査業務改善に関する調査研究事業 調査報告書（1）要約版資料の作成 統計作成ガイドブック」  
([https://www.meti.go.jp/meti\\_lib/report/2022FY/000670.pdf](https://www.meti.go.jp/meti_lib/report/2022FY/000670.pdf))
- 国連統計委員会 世界統計の日フォーラム 2025 講演資料  
神戸市「神戸市における EBPM の実践 ～日本の地方行政におけるビッグデータ活用～」p.17  
([https://www.soumu.go.jp/main\\_content/001043825.pdf](https://www.soumu.go.jp/main_content/001043825.pdf))
- 神戸大学大学院経営学研究科（分寺杏介）「統計的方法論特殊研究（多変量解析）講義資料 3 データの前処理」  
([https://www2.kobe-u.ac.jp/~bunji/files/lecture/MVA/html/chapters/03\\_data\\_preprocessing.html](https://www2.kobe-u.ac.jp/~bunji/files/lecture/MVA/html/chapters/03_data_preprocessing.html))
- 総務省統計局「なるほど統計学園」(<https://www.stat.go.jp/naruhodo/index.html>)
- 総務省統計局「品質評価事項チェックリスト」  
(<https://www.stat.go.jp/data/guide/pdf/checklist26.pdf>)
- 総務省統計局「結果表における機械判読可能なデータ作成に関する表示方法について」  
(<https://www.e-stat.go.jp/estat/html/machine-readable-stats-format.pdf>)
- 総務省統計局「データサイエンス・オンライン講座 特別セミナー 今日から始められる、Python によるビジネスデータ解析プログラミング講座」  
(<https://www.stat.go.jp/dss/pdf/programming/202412teach.pdf>)
- 千葉大学アカデミック・リンク・センター「データクリーニング入門（初心者向け）」  
(<https://alc.chiba-u.jp/eyr/2025/09/24/data-cleaning.html>)
- デジタル庁「オープンデータ取組の質評価指標（ベータ版）・研修資料  
オープンデータ研修テキスト 中級編（別添資料）データ利活用の参考ツール」  
([https://www.digital.go.jp/assets/contents/node/basic\\_page/field\\_ref\\_resources/651e3f74-9a8e-4797-98e9-97f17b0dcce7/343cf1bf/20250724\\_resources\\_open\\_data\\_materials-for-le.pdf](https://www.digital.go.jp/assets/contents/node/basic_page/field_ref_resources/651e3f74-9a8e-4797-98e9-97f17b0dcce7/343cf1bf/20250724_resources_open_data_materials-for-le.pdf))
- 東京大学数理・情報教育研究センター「数理・データサイエンス・AI 教育強化拠点コンソーシアム」  
([https://www.mi.u-tokyo.ac.jp/6university\\_consortium.html](https://www.mi.u-tokyo.ac.jp/6university_consortium.html))