

第2講 データの種類と収集方法

1. はじめに

教育研究において、データは主観を客観化し、経験を共有可能な知識に変換するための重要な基盤です。統計学は、教育現場で得られたデータから客観的な根拠を提供し、一般化可能な知見を得るための強力なツールとなります。

授業研究や実践において、主観的な感覚や学習効果を示すのは難しいものです。しかし、教育の品質を上げ、改善を進めていくためには、やはり客観的な根拠が必要となります。データを収集し活用することで、印象や経験則だけでなく、確か証拠に基づいた教育実践の改善が可能になります。

本講では、データの基本的な種類と特徴、そして教育現場における適切な収集方法について学びます。データを収集することは大変な作業ですが、それだけの価値があるということをまず押さえておいていただきたいと思います。

2. 教育研究におけるデータと統計の役割

2.1 教育研究におけるデータとは

教育研究におけるデータには、以下の3つの重要な役割があります。

(1) 主観を客観化する

教員の「感覚」や「印象」を数値や記録として表現できます。主観的な効果を客観的に見られるようにすることで、教育実践の効果を明確に示すことができます。

(2) 経験を共有可能な知識に変換する

個人の経験を他者と共有できる形にします。経験則を重ねていく中で、その経験を有益なものとして共有可能な知識に変換するためには、見直しや確認が必要です。データはそのための重要なツールとなります。

(3) 教育実践の改善サイクルを回す基盤

こうした蓄積がなされていくと、教育全体の改善、実践の改善をサイクルとして回せるようになります。データはそのための基盤を作る重要な要素です。

2.2 教育研究における統計の役割

統計学は教育研究において以下の3つの重要な役割を果たします。データが集まつくると統計的な分析が可能になり、様々な知見を得ることができます。

(1) 客観的な根拠を提供する

「この学習活動は効果がありそう」という印象を、数値で検証できます。個人の経験や直感だけでなく、データに基づいた判断ができるようになります。数値に振り回されることは避けたいところですが、数値的な評価が出てくると、客観的な根拠として使えるものに

なります。

(2) 一般化可能な知見を得る

目の前の 30 人のクラスでの結果が、他のクラスや学校でも当てはまるかを推測できます。特定の文脈での観察を超えて、より広い適用可能性を議論できます。統計処理をした数値化されたデータがあれば、他の授業や他の応用でも使える知見として育てることができます。

(3) 複雑な現象を整理・理解する

多くの要因が絡み合う教育現場で、何が重要な要因かを見極めることができます。データから意味のあるパターンを抽出したり、要因の相対的な影響力を比較したりできます。統計処理や解析ができるということは、推定をするといったようなことが可能になり、複雑な現象を整理・理解するための解析をかけやすくなります。

3. データの種類と特徴

3.1 情報とデータの関係

世の中は「物質」「エネルギー」「情報」の 3 つで構成されています。情報学では、この 3 つの組み合わせで世界を理解します。物があり、それはエネルギーがないと動きません。エネルギーを入れることによって動くことができますが、しっかりととした情報がないと非常に危険な動き、もしく

は予想外の動きになってしまい制御ができません。したがって、きちんと情報を入れて動かしていくということが一般的な考え方となります。この制御や正しく動かすという際に、情報データというものが非常に重要になってきます。

情報とは、文字・数字などの記号やシンボルの媒体によって伝達され、受け手に状況に対する知識や適切な判断を生じさせるものです。これについてデータ化してまとめていくというのが、今回の話題の中心になります。

3.2 記号化とデータ化の意味

たとえば、花を見たとき、人はそれを「パターン」として認識し、「花」という記号で表現し、さらに「種子植物の生殖器官」という知識として理解します。

人間の知的活動には、こういった情報を記号化して、他者や未来の子供たちに伝えるといったまとめ方をしているという前提があります。情報を長期間保存し、他者と共有するためには、このような記号化（デジタル化）が必要となります。

従来は口伝といったような形で人から人へ情報を伝えてきましたが、IT革命、情報化の波が押し寄せた結果、人が口伝するよりももっと確実にその事実を長期に残せるような技術が手に入っていきます。授業や教育活動の中でも、こういったICTの力を借りるのは非常に当たり前になってきています。

3.3 デジタルとアナログ

情報データというのは、いくつかの特徴によって保存のされ方が変わってきます。

(1) デジタル

- ・離散的に整数値で表現すること。デジタルコンピュータでは2進数（0と1）で表す
- ・正確な記録と再現が可能
- ・長期的に残しやすい
- ・離散的なため、中間の数値を抜いてしまうという特徴がある
- ・連続した滑らかな波ではなく、離散状態になり、情報が落ちる可能性がある

(2) アナログ

- ・連続した量を他の連続した量で表示すること
- ・時計や温度計などがその例で、あいまいな表現が可能
- ・中間のデータ、情報は全部残っている
- ・長期保存については、残すべきか残さないべきかという判断が必要
- ・冗長な表現や、場合によっては不要かもしれない情報が含まれる

このトレードオフをどう扱うかというところが情報学やコンピューターサイエンスでよく話されているところです。データ分析では、連続値（アナログ）を離散値（デジタル）に変換することで、コンピュータでの処理が可能になります。

3.4 量的データと質的データ

教育研究で扱うデータは、大きく「量的データ」と「質的データ」に分けられます。データの形式の話もありますが、その内容や特徴においての分け方もあります。

(1) 量的データ（数値化できるデータ）

量的データには以下のようないわゆる「数値化できるデータ」があります。

- ・テストスコア、出席率、課題提出数
- ・アンケートの評定値（5段階評価など）
- ・学習時間、発言回数

特徴：

- ・統計分析が可能
- ・客観的な評価ができる
- ・大規模調査に向いている
- ・数えることができる
- ・客観性が高く、調査もしやすい

注意点：テストの点が取れていれば全てを理解できているかというのではなく、必ずしもそれは限らない。量的データだけでは捉えきれない側面があることを認識しておく必要がある。

す。

(2) 質的データ（数値化が難しいデータ）

質的データには以下のようなものがあります。

- ・学生の自由記述回答
- ・インタビューの発話内容
- ・授業観察のフィールドノート

特徴：

- ・文脈や意味の深い理解が可能
- ・少数のデータでも深い洞察が得られる
- ・「なぜ」という問いに答えられる
- ・学習者の内的な状態や、教員がそのときに何を考えて話をしたかといった文脈、判断した深い意味の理解を表現できる

重要性：質的データを収集することはできても、それをどのように解釈するのか、分析するのかは非常に難しいものです。しかし、学習研究では、こういったデータを解釈して添えておくことが非常に重要になります。

(3) 両方のアプローチの重要性

現状としては、統計的な処理をした数値化されたデータと、質的なデータの両方を収集できるようになってきています。この両面から解釈するというのが、研究の中でごく自然になされています。

今後、収集するデータの種類が増えてくると、複数のアプローチから学習者の状況を推し量るということがますます増えてきます。そのため、データを収集する、蓄積する際に、どういったものを蓄積しておくべきなのかということは、ますます重要な議論になってきます。

3.5 データ駆動型処理と概念駆動型処理

人間の情報処理には 2 つのアプローチがあります。

- ・データ駆動型処理（客観科学的）：入力された情報の特徴を抽出し、パターンを認識する論理的な処理
- ・概念駆動型処理（主観科学的）：既有知識・経験・期待を基にした確認処理

教育研究では、これら 2 つのアプローチが相互に影響し合いながら、解釈、理解、学習が進んでいきます。

4. 尺度の種類

データを分析する際、データがどのような尺度（測定の水準）で測られているかを理解することが重要です。データを収集するときには、尺度と呼ばれるものがあり、代表的な 4 つが挙げられます。

す。

4.1 名義尺度（カテゴリー）

- ・単なる分類やラベルを表します
- ・どんなラベリングをするのかということが重要
- ・例：性別、専攻、出身地、学習スタイル
- ・使用できる統計：度数、割合、カイ二乗検定

4.2 順序尺度（順位）

- ・順序に意味がありますが、間隔は等しくありません
- ・順番があるものについては、きちんと順番も記録する必要があります
- ・例：成績（A・B・C）、理解度（理解できた～理解できない）、満足度（とても満足～全く満足しない）
- ・使用できる統計：中央値、順位相関

4.3 間隔尺度（等間隔）

- ・間隔が等しいですが、絶対的な0点がありません
- ・偏差値やばらつきというものを解釈するための感覚尺度
- ・例：偏差値、IQスコア、温度（摂氏）

- ・使用できる統計：平均値、標準偏差、相関係数、t検定

4.4 比率尺度（絶対0点あり）

- ・間隔が等しく、絶対的な0点があります（0 = 何もない状態）
- ・全体を推し量るために、比較する率で考える
- ・例：テスト得点、学習時間、年齢
- ・使用できる統計：すべての統計手法が使用可能、比率の計算も可能

これらの尺度をどう扱うかということについても、議論をしていく必要があります。

5. データの収集方法

教育研究においては、学習過程、どういった学び方をしているのか、何がつまずいているのか、もしくは早い段階で何か新しい推定に成功しているのかといった学習の内的な状況を見るために収集することが主となります。

学習過程や教材の提供タイミング、説明、学習活動が行われている際に発生する情報は、より質の高い学習を構築・検討するために非常に重要な情報となります。教育研究では、目的に応じて適切なデータ収集方法を選択することが重要です。

5.1 学習成果物・質問紙

収集対象：

- ・学習者自身が書き込んでいるもの
- ・読んでいるときにメモをした本
- ・レポート、課題、テスト
- ・プレゼンテーションの内容
- ・アンケート調査
- ・ポートフォリオ

特徴：

- ・学習過程を捉えるために有効
- ・学習成果を直接確認できる
- ・課題として提出する方法で収集可能
- ・最近ではタブレット教材も多く、タブレット上で作成されたログも含めて収集可能

注意点：

- ・ルーブリックを使った評価など、解釈方法が難しい場合がある
- ・質的データの場合、数値として推し量ることが難しい

- ・授業デザイン通りにしっかりと活用ができているかといった評価方法が必要

5.2 観察

収集対象：

- ・授業中の学習者の様子
- ・グループワークでの相互作用
- ・教室環境の影響

特徴：

- ・実際の行動を直接捉えられる
- ・学習者の実際の行動を直接見ることができる
- ・無意識の行動や相互作用を把握できる
- ・自己報告では得られない情報が得られる
- ・学習者自身が気づいていないけれども無意識に行っている学習活動や他者との議論の相互作用の状況を見ることができる
- ・自分では自認していないので報告はしないけれども、学習活動において重要な活動、動き、判断が見て取れる

注意点：

- ・観察者の主觀が入りやすい
- ・時間と労力がかかる
- ・観察されることによる行動の変化（ホーソン効果）
- ・観察の方法や学習研究に経験がある人でないと、なかなか拾うことができない
- ・あまりじっと見てもやりにくいので、カメラで撮る、気づかぬうちに学習者の学習を阻害しないように収集するといった配慮が必要

5.3 インタビュー

収集対象：

- ・学習者の考え方や感情の理解
- ・教育実践の背景にある意図の把握
- ・困難さや課題の詳細な把握

特徴：

- ・深い理解が得られる
- ・予想外の発見がある

- ・「なぜ」を探れる
- ・観察よりも直接学習者本人にいろいろ問うことができる
- ・半構造化インタビューなど、質問をすることはある程度決めているけれども、話の流れとして盛り上がった方についてもう少し深掘りをするといった臨機応変な対応が可能
- ・学習者が何を考えていたかという深い理解が得られる
- ・教員側が持っていたこと以外のことが発生していたという予想外の発見がある

注意点：

- ・時間がかかる
- ・分析が複雑
- ・インタビュアーのスキルに依存
- ・インタビューについての経験ややり方を知っておく必要がある
- ・満足にインタビューについてデータが取れたかは、人によって、場合によってムラができてしまう可能性がある
- ・データとして収集はできるけれども、扱いは慎重にしなくてはいけない

5.4 その他の既存データ

収集対象：

- ・成績データベース、出席記録
- ・LMS（学習管理システム）のログデータ
- ・図書館の利用記録

特徴：

- ・授業以外でも収集している情報が使える場合がある
- ・生徒の属性を知るためには必要

注意点：

- ・研究目的で収集していないものは注意が必要（研究倫理）
- ・データの目的外使用に関する倫理的配慮
- ・個人情報保護への配慮
- ・データの質と完全性の確認
- ・目的として研究目的で収集していない場合もあるので、研究倫理の配慮が必要
- ・取り扱いは簡単にできず、注意をして確認の上で使う必要がある
- ・場合によっては同意を得る必要がある

6. データ収集の実践例

いろいろな手法でデータを収集して、研究のため、よりよい授業のために分析・改善をしていきます。ここでは、実際の研究実践の例として、収集方法の具体例を示します。

6.1 協調学習場面における観察ツール（ROG）

協調学習場面における観察者の解釈を電子的に記録・共有する支援ツール（ROG）の例があります。このシステムでは、観察者用タブレット PC から赤外線通信でキャプチャデータを送信し、リポジトリに保存します。

これにより、学習過程を詳細に分析・解釈した経験から学生の状態を把握し、最適な学習を作る形成的評価の実現が可能になります。蓄積しておけば、後からこういった解釈・分析をすることもできます。

●大学授業における包括的データ収集の例

大学生を対象とした約 80 人から 100 名の授業に対して、学習活動、グループ活動を記録した例があります。当時（20 年ほど前）は、パソコンを授業中に使うというのは結構珍しい状況でしたが、以下のような取り組みを行いました。

背景：

- ・当時は紙や黒板での授業が当たり前だった時代
- ・ICT の力を借りて、学習者が必要なときに知識にアクセスし、他の人が考えているものについて

もオンライン越しでアクセスして議論を深めていくという斬新な新しい教育方法を目的として実施

収集したデータ：

- ・教案（word データ）：教員が事前に用意している講義案、時間配分・目標・話す内容などで構成
- ・配布物（word データ）：学生に配布する資料、作業用紙
- ・観察メモ（紙面）：講義中に教員・TA が作成する記録
- ・学生ノート（PDF）：講義中に学生が記入したノート
- ・ReCoNote 利用履歴（テキストデータ）：講義中に利用した学習システム履歴
- ・音声データ（mp3 など）：講義中の教員・学生の発話データ（全学生に IC レコーダーを装着し、誰がどの時点で何の話をしていたかを全て記録）
- ・映像データ（QT など）：講義風景の映像データ

システムの特徴：

- ・パソコンを間にて学生たちが資料を読んで説明し合う活動
- ・パソコン上でどういう動きがあったかを自動的に収集するシステム
- ・その場で教員が観察した結果をメモする機能
- ・これらすべてをポータルで見られるようにした

従来はパソコンでの蓄積がなかなかできなかつたものもデジタルデータとして収集できるようになりました。

6.2 学習環境と協調学習の効果の検討

インタラクティブな学習環境（MILAiS）と通常教室での協調学習の効果を比較した研究例では、以下のような知見が得られています。

- ・協調活動に適した学習環境の要件の解明
- ・アイデアや知識の外化→共有一再吟味の活動設計プロセスの詳細分析
- ・学習成果物の質的評価（文章としての成立、業務日誌としての体裁、書き方のポイント、自分なりの工夫）

普通の教室で議論をする場合よりも、アクティブラーニング型の机や話す体制でやった方が議論が活発化するということを可視化しました。こういった活動で最終的な成果物にも変化が出てきたりといったことで、活動プロセスというものをどうデータ化するかという例として参考になります。

これらのデータ収集と分析により、教育実践の改善に向けた具体的な示唆が得られます。

7. データの信頼性と妥当性

データを集めて、それに対して何らかの統計処理をする、もしくは解釈を与えるということは当然できます。データも昔よりは収集しやすくなり、気軽にデジタルツールを使って集めることができます。

しかし、その集めたデータ自体が正しいものなのか、その学習者を本当に表すものか、学習者の内的状況を推し量るために妥当なデータであるかどうかということは重要になってきます。

7.1 信頼性 (Reliability)

データの一貫性や再現性を示す指標です。

- ・再テスト信頼性：同じ測定を繰り返したときに同じ結果が得られるか
- ・一貫性：測定項目間で一貫した結果が得られるか
- ・評定者間信頼性：複数の評価者が同じ結果を出すか

7.2 妥当性 (Validity)

データが測定したいものを正しく測定しているかを示す指標です。

- ・内容的妥当性：測定内容が目的に適しているか
- ・基準関連妥当性：他の基準と適切に関連しているか
- ・構成概念妥当性：理論的な概念を適切に測定しているか

重要な問い合わせ：学習者の状況や過程を捉えるために適切なデータであるか？

7.3 データの質の確保

数値化されたものは特に、計算や統計的な処理をすれば、ある程度の評価値を出すことができま
す。しかし、その数値化されたものがどの程度信用できるものか、その学習者を表すのに妥当なも
のであるかというロジックがしっかりしていませんと、結果を出したとしてもあまり意味がない分
析となってしまいます。

学習者の状況や過程を捉えるために適切かどうかということは、収集段階から考えられる内容で
す。そういった意味でも、授業の設計・デザインをしっかりする必要があるということに立ち返っ
ていただけるといいと思います。

7.4 倫理的配慮

データの種類のところでも述べましたが、もともとその授業をより良くする、学習研究を目的に
していないで集めているデータを使うことがある場合があります。その場合は、やはり個人情報等
を含む内容ですので、適切な処理が必要です。

具体的には：

- ・ 同意を得る
- ・ この情報は収集したら授業改善に使う可能性があるという断りを入れる
- ・ データを使うときには倫理的な配慮が必要

今後ますます重要になってきますので、データを収集するときには、こういった同意を得る、使用者に対して何に使うかということをしっかりと説明するということを心がけていただきたいと思います。

8. 良いデータ収集のポイント

教育研究において良質なデータを収集するためには、以下のポイントを押さえることが重要です。せっかく集めたデータを分析なり解釈なりしていくわけですが、やはり集めたデータで何か算出するというのは当然できますが、教育研究においては以下の点が重要です。

8.1 研究目的に合った方法を選ぶ

- ・研究の問いは何か？
- ・どのような情報が必要か？
- ・量的データと質的データのどちらが適切か？
- ・その目的に合ったデータが適切な方法で収集されているか

8.2 複数の方法を組み合わせる

- ・異なる視点からデータを収集
- ・結果の信頼性を高める
- ・より豊かな理解を得る

8.3 限界を認識し、解釈に反映させる

- ・どのようなデータ収集方法にも限界がある
- ・バイアスの可能性を認識する
- ・結果の解釈に慎重さを持つ
- ・学習過程を推し量るには、データを集めたところで、その人の学習者の内部で起こっていることをすべて書き出すというのはほとんど難しい
- ・収集したデータでは、この内容はわかるけれども、ここが限界だということで、限界点をしっかりと押さえた上で、誠実に解釈をして、その学習成果を報告していくという考え方方が重要

8.4 実現可能性を考慮する

- ・時間：データ収集と分析にかかる時間
- ・コスト：必要な費用
- ・人員：必要な人的リソース
- ・倫理：研究倫理審査の必要性

9. まとめ

本講では、教育研究に主に焦点を当てて、データがどのような使われ方をするべきかという話から始め、その種類、特徴、集めたときにどういった整理方法があるか、信頼性と妥当性、またデータ

タを使うときには倫理的な配慮が必要になってくることについて触れました。

重要ポイント：

- ・データの種類　：デジタル・アナログ、量的データ・質的データの違い
- ・尺度の理解　　：名義尺度、順序尺度、間隔尺度、比率尺度の特徴と統計手法
- ・データ収集方法：学習成果物・質問紙、観察、インタビュー、既存データなど、目的と用途に応じた選択が重要
- ・信頼性と妥当性：データの質を保証するための重要な概念です
- ・倫理的配慮　　：研究倫理と個人情報保護への配慮が必要です

教育研究においてデータを適切に収集・活用することで、主観的な経験を客観的な知識に変換し、教育実践の改善につなげることができます。データを収集することは大変な作業ですが、やる価値があるということを押さえておいていただきたいと思います。

次講以降では、収集したデータをどのように前処理し、分析していくかについて学んでいきます。

参考文献

- ・総務省統計局、レベル別テキスト、<https://www.stat.go.jp/dss/getting/textbook.html>
- ・総務省統計局、なるほど統計学園、<https://www.stat.go.jp/naruhodo/index.html>
- ・ソコスト、<https://soco-st.com/>

- ・中溝幸夫, 箱田裕司, & 近藤倫明. (1983). リンゼイ/ノーマン情報処理心理学入門 I—感覚と知覚. P.8-16
- ・平成 11 年度 CREST 「高度メディア社会の生活情報技術」研究領域 三宅なほみ「高度メディア社会のための協調的学習支援システム」