

教師のためのデータ分析方法 I

Data Analysis for Teachers I

石川 紀宏

目次

はじめに	84
I. 従来の実践報告の批判的検討	84
II. 記述統計によるデータ分析方法	84
III. 推測統計によるデータ分析方法	86
おわりに	89
参考文献	90

はじめに

今後における日本の学校教育のあり方を熟考したとき、特別支援教育を含めて言えることは、根拠のある実証性の高い結果を要請する潮流が眼前に迫っていると思う。Evidence-based practice という潮流である。学校現場で教育実践をしている教師のひとりとして、筆者はどのようにその要請に応えるのか、ひとつは自らが実践してきた実績から応えたと思う。本校の過去10年ほどの研究紀要に執筆し研究成果として示してきたからである。その背景には、学習主体である子どもに学習の成果として確実に身につけたことを、学校教育における指導の主体である教師として評価しなければならなかったからである。(参考文献、参照)

そのような経過を経て、筆者が提示する「教師のためのデータ分析方法」とは、どういう意味か明示しておきたい。それは、教師が日々の教育実践から入手したデータを分析して、自らの実践を反省し、その指導の効果をより実証的に、より科学的に判断するために遂行する技法のことである。簡略的に言えば、「統計的方法(記述統計と推測統計)」あるいは「量的方法」となるが、何故、そのような方法を提起するのか、それは、以下の5点の動機にある。

- ・自分の教育実践に自信と根拠を持ちたい。
- ・子どもの学習の成果を子どもに示したい。
- ・保護者への責務として示したい。
- ・当校の成果として示したい。
- ・若い教師の世代に伝えたい。

今回は、特に、最後の「若い教師の世代に伝えたい」がために本研究を遂行することにしたものである。

統計的方法を学ぶ教師は非常に多い。しかし、その多くが途中で挫折している。そのような教師たちに理解できるように、簡便な方法に限定し、公刊された実践報告を統計的方法の視点から批判的に検討しながら、より有益なデータ分析の手法を開示することにする。

I. 従来の実践報告の批判的検討

ここで取り上げる批判的検討の対象は、多くの実践報告に共通する問題点でもある。(文献15, p187-193. 参照)

- ・対象児童のアセスメントが欠落している点
- ・学習評価の基準が不明確である点
- ・学習回数の記述がない点
- ・学習効果を判定できない点

まず、第一に、アセスメントの欠落について、今日の教育研究レベルでは、対象児のアセスメントは不可欠である。例えば、本研究紀要に執筆されている昨年度の当校の研究協議会第1分科会算数・数学の報告を参照されたい。図形の認知と筆記に課題のある子どもに対

して WISC- III の検査結果、「視覚-運動統合発達検査」(VMI) の結果、「フロスティック視知覚発達検査」の結果を利用している。通常このようなタイプの研究では、指導前と指導後の検査結果を比較するデザイン、いわゆる「プリ・ポスト研究デザイン」を活用する。(注: 専門的には、いくつかの弱点がある研究デザインではあるが、ここでは言及しない。) ここで重要な点は、対象児の知的レベルと視知覚の発達レベルが不明であり、視知覚の問題ではなく、単純に学習経験不足のため書写ができない可能性がある。

第二に、「単元の評価規準」と「個々の判定基準」は明記されているが、使用した PLM 知覚-運動学習教材とその指導方法の手順や動機づけが述べられているだけで、子どもの学習の経過と対応した記述が十分ではない。すなわち、第三に、学習回数の記述がなく、実際どのようなデータをどの程度収集し分析したのか不明である。単なる主観的判断であると実践報告を読む教師からは見られるであろう。そのため事例の写真があってもどのように読み取ればよいのか不明である。実践記録や研究報告を読む教師たちは、実際の指導を観察してないのであるから、読み手の教師が客観的に判断でき追試可能な検査結果、データ、表、グラフなどの資料を用意する必要がある。

以上の各点を考慮して言えることは、そのような実践報告では、実証性もなく客観的根拠もないので学習効果を判定することはできない。それ故に、実践の指導効果があったという主張はできないであろう。では、どのように実証性があり客観的根拠のある実践を提示できるのか、以下の行論で示していきたい。

II. 記述統計によるデータ分析方法

具体的にどのようにすれば、より実証性があり、より客観的な根拠のある実践報告となるのか、そして、学習効果があると示唆できるかについて報告する。

まず、確認したいのは、私たち教師がなんらかの指導を遂行し子どもが学習し、成長し発達したといえるのはどのような現象が生じたときであろうか。初期状態 A から次の状態 B に変化し、なおかつ、その変化が有意味なとき、価値を持つと判断されたときわれわれは、成長した、発達した、伸びた、広がった、深まったなどの記述ができるのである。そのような記述ができるように検査をしたり、調査をしたり、実験をしたり、学習の記録を取ったりしてデータを収集し分析して、他者にも分かるように判定を下すのである。

さて、以下に述べるのは、個人レベルの事例であることを最初にお断りしておく。個別の指導計画の作成が「義務化」されてから、障害児の指導にあたっては一人ひとりに対応することは当然視されてきており、個々の成長発達が課題とされているからである。無論、小集

団で指導するケースも多いのであるが、今回は個人レベルに限定して例示していきたい。統計的方法では、まず、記述統計の方法を利用することにする。しかも、教師一人で実行できる方法であることが、その特徴である。

1. 事前・事後、プリ・ポストデザインの場合

前述したが、この方法は標準化された検査を事前事後に実施して、指導の効果を判定する方法であり、教育現場では多用されている。例えば、視知覚の検査や視覚を必要とする認知課題を利用した指導を想定してみよう。図形を描くことに困難を伴う子どもがいるとする。その子どもの事前検査として「視覚-運動統合発達検査」(VMI)や「フロスティック視知覚発達検査」を実施しておき生活年齢7歳3ヶ月、VIM年齢4歳4ヶ月の肢体不自由の子どもが、ある種の認知課題を伴う図形描写を3ヶ月程度学習したとする。(自立活動の時間の指導

毎週1回 祝日や欠席を考慮して 12回から15回程度学習回数が確保されたとする。)最終日に、同一検査をしたら生活年齢7歳6ヶ月、VIM年齢5歳3ヶ月に向上していたとする。結果は、3ヶ月分の教育的諸要因と成熟要因を考慮しても図形描写の認知的、技能的課題の学習効果は明瞭に示唆できるものである。結果の表示方法としては、表にしてもよいし、グラフ表示にしてもよい。自立活動においては、技能や身体的な訓練的要素が強い内容を指導する場合や教科的要素を伴う学習には、少なくとも3ヶ月程度の指導期間が必要である。

注意すべき点は、どのような学習や訓練でも技能的あるいは教科的な要素の強い学習では、その効果は継続的な指導と学習回数に伴いゆるやかに変化するものであり、応用行動分析などが遂行する逸脱行動の変容の事例が示す、適正な要因を伴う介入により急激な変化として発現することはないものである。また、現場の教師は重度・重複障害児の指導においても、同様に学習効果は非常にゆるやかに変化することを経験知としても知っている。

2. 指導の前半と後半を比較する場合

前述の方法は、標準化された検査を利用して指導前と指導後と比較して、結果を示し指導効果があったと判定するものである。しかし、いつでも標準化した検査が適切に使用できるとは限らない。検査がなかったり、検査方法を知らなかったりすることはあるものだ。そのような場合、指導の前半と後半に二分割して比較する方法を実行すると良いだろう。

この場合、二通りの傾向がある。まず、上昇傾向を示すことが有意である場合、学習回数に比例して増加傾向を示すようなデータを収集している場合である。例えば、1分間に読める漢字の数、3分間に書き出せる英語の不規則変化動詞の数、5分間に計算できる加減算の数などである。結果の例示は、通常、前半と後半の同回数の平均値や中央値を算出して比較する。ただし、実施回

数が奇数回の時は、真ん中の数値は取り除き、前後とも同数にしてから処理する。前半よりも後半の方が増加していることを表やグラフで表示すればよい。(図1 参照)

次に、下降傾向を示すことが有意である場合、学習回数に反比例して減少傾向を示すようなデータを収集している場合である。車椅子で100メートルを自力走行する時間、算数の100マス計算問題の平均解答時間、速いほど数値は小さい値を得る。基準を明確にしてデータ収集し比較すればよいのである。ただし、100マス計算問題の例は、正確さと速さの2つの基準が必要になり、両者の関係は負の相関を示すことになるが。結果の例示は、前者の増加傾向の場合と同様に数値処理をしてから、前半よりも後半の方が減少していることを表やグラフで表示すればよい。

このデータ分析方法の利点は、データ数が少なくても前半と後半に二分割して比較できる点にある。実際、5回程度の学習回数しかない授業の場合も利用できるからである。長期欠席のケース、出席できても発作や体調不良で授業にならないケース、手術後のため実施できなかったケースなどの時、活用できるのである。

3. 全データを利用する場合

授業回数が確保でき、学習回数が多く比較的多くのデータが収集できた場合の分析方法について紹介する。この方法は、前半と後半のように二分割して比較するのではなく、全データから傾向や特徴を明確にする方法である。前述のように上昇、増加傾向を判定したり、下降、減少傾向を判定したり、特別な場合には安定傾向を判定したりするものである。

国語や算数などの教科的学習の中で、基礎・基本となる学習内容を定着させ習熟させる学習は、毎日取り組むことが多く、比較的学習回数が多くなりデータ数が多い。漢字の読み書き5分間テスト、10問からなる計算小テスト、自立活動で毎回取り組む上肢の基本訓練や手遊び活動などを想定すればよい。それらのデータを時系列に配列して、グラフ化すれば傾向は明白になる。この場合、教師の指導による学習回数が独立変数扱いとなり、子どもが学習しその結果を計測したデータが従属変数扱いとなる。

そのようなデータはグラフ化してみると、通常、次のようなタイプに分かれることが知られている。

- ・単調増加傾向を示すタイプ
- ・単調減少傾向を示すタイプ
- ・上下に変動するタイプ
- ・ほとんど変化しない安定タイプ
- ・天井効果を示すタイプ
- ・フロアー効果を示すタイプ

これらのタイプで説明が必要なのは、ひとつが上下に変動するタイプである。このタイプは、学習にむらがあ

り集中力に欠ける子どもに多い。いまひとつは、ほとんど変化しないタイプである。このタイプは、努力はしているがその学習に限界を示している子どもに多い。前者の子どもには、その子どもでも集中できる程度に課題のレベルを落とすなり、課題を分割する、あるいは、スモールステップ化することである。後者の子どもには、課題の難易度を落とすか、下位の課題に習熟させてから再度取り組ませる必要がある。

さて、入手できたデータ数が比較的多ければ ($n > 30$) EXCELのグラフ機能を利用して折れ線グラフで表示すればよい。X軸に学習回数を取り、Y軸にデータ数値を取り表示すれば変化の様子が全体として把握できる。グラフ化したらグラフの折れ線をクリックすると直線の当てはめができる機能がある。一次直線を選択すると最小自乗法による直線の当てはめが自動的に表示されるので、従属変数の全体的な傾向性が明らかになる。(図2. 参照)

4. 一致率を利用した場合

記述統計を利用することに関して、今までは教師一人だけで実行できるデータ分析の方法について言及してきたが、最後に二名の教師による判定方法のひとつ、一致率を利用した方法を紹介しておきたい。特別支援教育では、教師間の共通理解を促進することが話題として取り上げられることが多いので参考になるだろう。

最も簡単な場合で考えてみたい。二名の観察者(教師)がある教育現象をそれぞれ独立に判断する場合である。まず、大前提は観察の独立性を確保することである。他者に相談しないで、自分の公正なる判断(教師の内的な価値基準に基づくこと)で観察項目について、Yes/No、○×、3段階、5段階評定などを実施することである。事前に相談していたら一致率は当然上昇するので不公平となるから、注意したい。

計算式は、次の通りである。

$$\frac{(\text{総項目数} - \text{不一致項目数})}{(\text{総項目数})} \times 100 =$$

表1 教師二名による授業評価

	A	B	C	D	E	F	G	H	I	J
n	○	○	×	○	○	○	×	○	○	○
m	○	○	○	○	○	○	×	×	○	○

今、nとmという教師二名がAからJの10評価項目により、ある授業を観察して評定したとする。上記の表のような結果が得られた。一致率を計算すると次のようになる。一致した項目数8、不一致項目数2、総項目数10。 $8/10 \times 100 = 80$ 一致率は80%となるが、G項目は×で一致しているので、○で一致しているのが問題であるか

らその分減じて一致率は70%とすべきである。確実に教師二名の判定が一致しているといえるには、90%以上の一致率が必要である。最低でも80%以上は必要である。(だから、評価項目が20項目ならば17項目以上が一致する必要がある、30項目ならば25項目以上一致する必要がある。)なお、評定が異なったCとH項目は、何故異なるのかお互い相談してみる必要があるだろう。

Ⅲ. 推測統計によるデータ分析方法

一般に推測統計は、母集団を決め、その母集団からランダムに標本を抜き出し、その標本に対して仮説を検証するために、ランダムに条件や要因を割り付け実験し調査する方法であり、検定の結果に基づき標本から母集団への推論を展開するものである。通常、統計的方法(仮説-検定)は次のような手順を含んでいる。

- (1) 問題意識
- (2) 仮説の設定
- (3) 統計モデル
- (4) 有意水準 (α)
- (5) 標本分布
- (6) 棄却域
- (7) 決定
- (8) 考察

このような統計的方法は、特定の集団・群間の比較実験や調査に採用されるものであり、論究してきた個人レベルのN=1デザイン、単一被験者計画などには適さない場合がある。なぜなら、母集団を仮定しないことがあり、母集団への推論や法則定立的な一般化を目的としないからである。仮に母集団を特定しても、極めて小さい母集団の場合があるからである。例えば、脳性まひの痙直型の小学校低学年児を対象にしたとする。障害児の集団に含まれる肢体不自由児の集団から脳性まひの痙直型の集団のうち低学年児となる集団は、健全な低学年児の母集団と比較したら極めて小さな母集団である。

現場の教師の立場からすれば、母集団への推論や法則定立的な一般化の研究は、主に専門家である大学や研究所の研究者の遂行課題として考え、「学問の自由」の立場では同等であるが、特別支援教育の立場では、データから有益な情報を引き出すような、つまり、特定の指導方法や技法が特定の子どもの学習に効果があるか、評価に活用できるか、という視点が重要であり、その方法が測定や計測の信頼性が確保でき、導き出された結果が現実に妥当性が高いことを示せるほど客観性や実証性をともなっているか、という視点が重要である。

それでは、記述統計の事例で扱ったそれぞれの場合に対応して、推測統計の事例を示していこう。

1. 事前-事後、プリ・ポストデザインの場合

ある教科のある単元について個別指導を実施して、事

前テストの得点と5回の個別指導を実施した後に、事前テストと同一の事後テストをしたとする。テストの問題数は15あるとする。あるいは、ある授業を事前と事後に参観してもらい、授業に関する15項目についてそれぞれ7段階評定をしてもらったとする。このような事例があったとして、入手できるデータは、前者はテスト得点、後者は評定点である。前者の場合で説明すれば、「帰無仮説は、指導の前後の成績は変わらない」「対立仮説は、指導後の成績は伸びている」となる。この仮説を検定するのに利用するのは、計測値であるテスト得点である。テスト得点が比較して、事前テストより事後テストのほうが高いという結論が出れば、成績は伸びたと判断できる。ただし、その単元の学習についてのみ言及できる。その教科全体に対しては言及できない。

実際には、データの尺度水準（名義尺度、順位尺度、間隔尺度）とデータ数（標本サイズ）、それに独立標本か関連標本（対応のある標本）により、利用できる検定群が決まってくる。データの尺度水準が名義尺度と順位尺度の場合、ノンパラメトリック検定を利用することになり、間隔尺度以上の場合、パラメトリック検定が利用できる。

無論、データが正規分布していること、分散が等しいことが仮定できれば、パラメトリック検定（ $n > 30$ ）が利用できるが、そうでなければノンパラメトリック検定を利用することになる。この事例だとデータ数が15で少ないので、データの尺度に関わらず、正規分布と等分散の仮定はクリアできそうになく、仮定に抵触するのでノンパラメトリック検定となる。ただし、明らかに正規分布していることが判明している母集団から、ランダムにデータ（標本）を10抜き取った場合、間隔尺度以上のデータならば問題なくパラメトリック検定が利用できる。

検定の具体例を提示しよう。ここでのデータは、ある教師の授業を観察している評定者が、授業に関する評定項目に基づき、事前事後の2回同一の7段階評定（注：評定の低い方から高い方へと数値を割り付けるのが基本である。）を実施しているとしよう。この事例では、データに対応があり関連しているので「関連標本」あるいは「対応のある標本」と呼ばれる。まとめたものが表2である。このデータを名義尺度と考えると、 χ^2 検定（マクネイマーの検定）、二項検定（サイン検定）、フィッシャーの直接確率計算が利用できる。順位尺度と考えるとウィルコクソンの符号化順位検定、並べ替え検定が利用できる。間隔尺度と考えると正規性と等分散の仮定に問題が残るが、対応のあるt検定が利用でき、限られた範囲の計測値しか取らない場合には、スネデッカー・コクランの連続性の補正をしたtc検定が利用できる。

多様な検定が利用できるが、最適な検定は検定力の強さから見て、この事例だとフィッシャーの直接確率計算か、ウィルコクソンの符号化順位検定がよいであろう。前者は超幾何学分布という特殊な分布を利用し、後者は

符号と順位を利用する検定である。前者を利用する場合、表から次のような集計表に変換する必要がある。つまり、事後の評定が高くなった数（評定の差がマイナスのもの）と低くなった数（評定の差がプラスのもの）に集計し直すのである。ただし、評定が同じものは除く（評定の差が0のもの）。

集計表

事後の評定	高い	低い	合計
	11	2	13

表2 授業の事前—事後の評価（7段階評定）

項目	事前の評定	事後の評定	評定の差
1	5	7	-2
2	5	6	-1
3	4	2	2
4	5	6	-1
5	6	6	0
6	4	5	-1
7	6	5	1
8	2	5	-3
9	3	4	-1
10	6	6	0
11	3	4	-1
12	2	4	-2
13	2	4	-2
14	6	7	-1
15	3	4	-1

フィッシャーの直接確率計算を実施したところ、 $n = 13, p = .0225$ （両側検定）。ウィルコクソンの符号化順位検定（ $n = 13$ ）を実施したところ $R = 15, p = .0317$ （両側検定）（注：5%よりも確率が低い）となり事後の評定が高いことが判明した。この授業の評価は、授業の観察者によると授業の改善を示唆している、とか、教師の指導方法に改善があったといえる、とか解釈することが許されるだろう。ただし、学習者である子どもの学習に効果があったという主張はできない。何故なら、間接的には子どもの学習活動に対して有益であると示唆することはできるだろうが、子どもの学習活動を調べているのではなく、授業に関する15評定項目により授業を観察者が評定しているからである。

2. 指導の前半と後半を比較する場合

授業を実施していて、前後2回の評定だけで指導効果を判断することには不安を感じる教師も多いのではないだろうか。そのような場合、何回も同様の学習を繰り返して定着させ習熟させることにより判断を下したいと考えるものである。そのような学習は、教科の基礎・基本に関係するドリル学習や反復学習の場合が多い。例えば、5分間の漢字読み書きテスト得点、10分間の算数の計算

テスト得点, 50m走のタイム(秒)などが想定できる。

子どもの学習回数が前半6回, 後半6回以上確保できれば, 最も簡便な2項検定の特殊ケース(事前確率が50%の場合である)である符号検定(サイン検定)が利用できる。標本サイズは大きい方が望ましいので, ノンパラメトリック検定ではできれば10回ずつ以上は欲しいところであるが, 最低6回から検定できるという意味である。以下に架空のデータであるが, 漢字の書き取りテスト10題中の正答数を示しこのデータを利用して学習効果があるか判定してみたい。

回数	1	2	3	4	5	6	7	8	9	10	11	12
正答数	4	5	5	6	6	6	5	6	6	7	8	9

前半の6回分と後半の6回分を対応(注: 1回目と7回目が対応するようにし, 最後は6回目と12回目が対応する。)するように並べ替え, 差を取る。符号がプラスかマイナスになるので, その数を集計する。同数値の場合は0になるので, 取り除いておく。

前半6回	4	5	5	6	6	6
後半6回	5	6	6	7	8	9
差の符号	-1	-1	-1	-1	-2	-3

集計表

符号の分類	プラス	マイナス	合計
	0	6	6

符号検定の結果は, $n = 6, p = .0313$ (両側検定)となりマイナスの符号が有意に多いことを示している。符号検定は, 絶対値の大きさは無視して符号だけを取り扱うので, 集計するとマイナスが6個プラスが0個の場合に分かれていることが分かる。マイナスが多いということは, 後半の数値がみな前半より大きいことを意味している。後半の数値の方が大きいことは, それだけ漢字の正答数が多くなっていることを示唆している。前半より後半のほうが漢字の書き取り学習が, 向上していることが判明した分けである。記述統計的に処理して, 前半6回の平均値5.33, 後半6回の平均値6.83として後半の平均値が高いので漢字の書く力が伸びたと判断しても, 無論かまわないと思う。ただし, この方法は, 前半と後半を同数ずつ対にして比較するので, 偶数ならば良いが13回分のデータがあるときなどは, 中央の7回目の数値は取り扱わずに除き, 前半6回後半6回にして算出する必要がある。13回目を除いてもよい。いずれにしても入手したデータの一部(符号のみ)しか利用しないので, 情報量の損出が多く検定力は弱い検定であるが, 初心者にも分かりやすく非常に早く簡便に利用できる点優れた検定であるともいえよう。

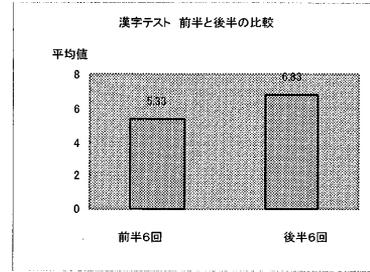


図1 漢字テスト 前半と後半の比較

3. 全データを利用する場合

前半, 後半に2分割してデータを処理するよりも, 入手したデータを全部利用して検定できないものだろうか。それは通常傾向検定と呼ばれるもので, 何通りかのものであるが, 使いやすく検定力のすぐれたMannの傾向検定がある。学習回数を独立変数として子どもの学習活動を計測した数値が従属変数となるものであり, ケンドールの順位相関係数 τ (タウ)を利用する統計的方法である。順位相関係数が正の符号ならば, 上昇, 増加傾向にあることを示唆し, 負の符号ならば下降, 減少傾向を示唆するものであり, その強さは相関係数の絶対値で表現される。 $(-1 \leq \tau \leq 1)$ 具体的な方法は, 計測値に対して順位に従い数値を割り振っていくのであるが, ただし, 説明は煩雑になるので省略したい。EXCELの統計関数にもあるので利用は簡単である。また, スピアマンの順位相関係数もあるが, ピアソンの積率相関係数を機械的に順位データに当てはめたものであり, 原理的には, 順位だけを扱うケンドールの順位相関係数が正しく,

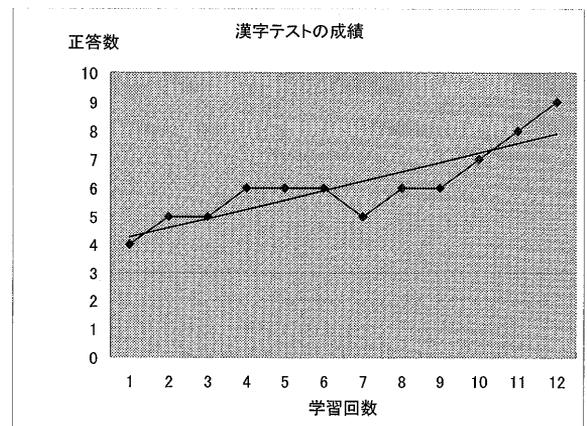


図2 漢字テストの成績

標本サイズが, 大きくなるに従い正規分布に接近するのがはやいという優れた特性がある

具体例は, 前述の漢字書き取りテストのデータを利用して順位相関係数を算出すると, $n = 12, \tau = .795, p = .0002$ (両側検定)となり正の高い相関を示唆している。検定の結果, 漢字テストの書き取りの正答数は, 学

習回数を経るに従い上昇傾向を示していることから、このテストでは漢字の書き取りの力が伸びていると判断できる。ここでは、上昇傾向を示すか、下降傾向を示すか、検定前に仮定していないので、慎重に両側検定を採用している。無論、上昇傾向が強く仮定できるのであれば、右片側検定で構わない。この方法の利点は、全データを使い全体の傾向を見ることができ、初期の状態からの変化過程がグラフ化することで視覚的にも判断しやすくなることである。符号も絶対値も利用するので情報量の損出が少なく、比較的簡便に利用でき検定力もパラメトリックな検定と比較しても遜色がない点にある。

4. 一致率を利用した場合

教師二名の判断が一致しているか否か、一致率を算出することで、記述統計的に説明したが、ここでは推測統計における複数の観測者あるいは評定者の判断に関する一致性あるいは一意性の検定について説明しておきたい。自分の授業記録をビデオに録画し、複数の教師に評価してもらうことは教育現場では日常的に行われていることである。特に、研究授業として厳しい批判にさらされることは、教師ならばだれもが経験することであり、授業力を高めたり、教師の力量を高めたりするために必要なことと理解されている。だが、多くの場合、先輩教師やベテラン教師の客観的根拠に欠ける主観性の強い批評をもらうだけで、反論することもできず悔しい思いだけが残ることもある。そのような時、予め用意された評価項目に基づき複数の評定者である教師が相互に独立に観察し評定した結果を参照しながら、討議したらより生産的になり、批評される教師にも納得しやすく優れた点、劣っている点、改善点などが明白になるのではないかと、しかも多くの判断が一致しているならば、そこには客観性の高い共通の意義が見いだされるのではないだろうか。このように評価項目を用意していても外的に適正な基準が構成できないとき、複数の評定者（教師）の内的基準に基づく判断の一致度を考えるのは合理的である。

まず、授業に対して、子どもの学習活動に対して、指導方法に対して、教材に対してなど2名の評定者の判断が一致している否かは、計測される項目が、○、×、Yes、Noなどの2値的なものであった場合と順位尺度や3段階5段階の評定である場合によって、検定方法が異なる。前者の場合 κ （カッパ）検定を使う。この検定は複数の評定者の場合にも拡張できる。後者の場合ケンドールの順位相関係数を使うことになる。評定者の数が k 人（3名以上）の場合は、入手するデータが名義尺度の場合で関連標本である時は、コックランのQ検定が使用でき、順位尺度の場合で関連標本の時は、ケンドールの一致係数 W カフリードマンの検定 F_r を使用することになる。この検定同士には直線的な関係があり相互に変換可能である。特殊なケースでは、複数の評定者にいくつかの対象について対比較させて、どちらがより好まし

いか評定させ、その一致度を調べるケンドールの一意性検定 u もある。

ここでは、比較的理解しやすい検定であるからケンドールの一致係数 W を紹介する。3名の教師（X Y Z）による10対象（評定したい項目A～J）に順位をつけ、その順位間の相関が有意であるかを検定する。架空のデータは以下ようになるが、実際には甲乙付け難く同順位（タイ）が生ずることがある。

	A	B	C	D	E	F	G	H	I	J
X	1	2	3	4	5	6	7	8	9	10
Y	3	4	1	2	8	5	9	10	6	7
Z	2	3	6	1	4	5	8	7	10	9

- ① 6 9 10 7 17 16 24 25 25 26
 ② 36 81 100 49 289 256 576 625 625 676

①が各順位和であり、②が各順位和の自乗値であり、一致係数 W を求める公式は以下の通りである。

$$W = \frac{12 \sum Ri^2 - 3k^2N(N+1)^2}{k^2N(N^2-1)}$$

k = 評定者数 (3) N = 評定対象 (10) Ri = 順位和の値、 $\sum Ri^2$ は順位和を自乗した数値の総和である。 W を求めると

$$W = \frac{12 \times 3313 - 3 \times 9 \times 10 (10+1)^2}{9 \times 10 \times (10^2-1)} = \frac{7086}{8910} = 0.795 \quad (df = 9, p = .011)$$

となり、有意水準5%以下となるので、教師3名による当該授業に対する10項目の評定の相関は有意であった。すなわち、評価項目による授業に対する判断は、3人の教師間で一致していたといえる。

おわりに

教師のためのデータ分析方法について、若い教師（教職10年以内）を想定して、比較的簡便な記述統計と推測統計の手法を開示してきたが、統計的方法の誤解について言及しておきたい。「データは集めたけれど、分析方法が分からない。」と統計的方法について尋ねる教師がいる。典型的な誤った研究の考え方である。仮説を設定しその仮説を検証するためにデータを集め分析するのであって、データを収集する前に分析方法を研究全体の計画の一部、内的手続きとして確定していなければならない。集めたデータを分析するのが統計的方法であると誤解しているのであり、初心者の若い教師が陥る隘路である。そのようなことにならないように、データ分析の事例をいくつか示してきた。データを分析するには、入手するデータの尺度水準（名義、順位、間隔）、データの次元（1次元、2次元、多次元）、データ数（標本サイ

ズ), 標本数 (1 標本, 2 標本, k 標本), 標本間の関係 (独立標本, 関連標本) などを調査・実験前に決めてから, それぞれに対応した検定方法を選択しておくのである。それは, 「習うより慣れる。」の世界かもしれない。身近にいる統計的方法による実践を遂行したことがある教師や研究者に相談しながら, 自分で指導仮説を設定し, 工夫した教材や指導方法を駆使して, 実際にデータを集め分析するという経験を積み重ねながら, 体得せざるをえないのである。

そのようなデータ分析の結果, 指導の効果が客観的に明白になったならば, 次に, 教師としてその成果を整理して, 文章化し実践報告, 実践研究としてまとめ上げて公表することである。できれば, 事前に統計の分かる教師や研究者に読んでもらい示唆や訂正を受けることができれば幸運である。データ分析の方法は, 最初に筆者が5点の動機を明示したが, 若い世代の教師たちにそれぞれの現場で実際に活用してもらわなければ, 伝達・伝承されたことにならないし有意義ではない。統計的検定により偶然では起こりえないことが, 自らの実践において生じたのであれば, それは偶然ではなく必然であり, 教師の指導の効果が明白になったことを意味する。それは大きな自信となり教師の確信となる。そのような努力を惜しまない若い教師たちの参考になれば, 筆者としては幸いである。

なお, データ分析方法に焦点を当てて論述してきた関係上, 研究論文の記述のしかたについては, ほとんど言及していない。近年研究論文の書き方について出版されることが多くなってきたものの, その方面に関しては, 20年前から優れたテキストとして, 既に田中敏らにより公刊されている著作があるので参照していただきたい。(文献16. 参照)

付記

本研究は, 平成22年度筑波大学公開講座『教師のためのデータ分析方法』の一部として記述されたことを明記する。

～参考文献～

1. 石川紀宏 (1998) 文学教育への統計的アプローチ～詩の指導効果を中心に～. 筑波大学附属桐が丘養護学校研究紀要, 第34巻, 57-66.
2. 石川紀宏 (1999) 重度・重複障害児への統計的アプローチによる遊具を媒介とした個別指導の効果. 筑波大学附属桐が丘養護学校研究紀要, 第35巻, 64-71.
3. 石川紀宏 (2000) 短期在籍児童のための運動学習の個別評価～ノンパラメトリック統計学による3分間走の分析を中心に～. 筑波大学附属桐が丘養護学校研究紀要, 第36巻, 58-66.
4. 石川紀宏 (2000) 個別指導計画に関する評価研究～複数教師による評価とその一致性～. 筑波大学附属桐が丘養護学校研究紀要, 第36巻, 41-48.
5. 石川紀宏 (2001) 詩の授業に関する評価研究～複数教師による評価とその一致性2～. 筑波大学附属桐が丘養護学校研究紀要, 第37巻, 39-48.
6. 石川紀宏 (2001) 個別運動学習の効果～下肢の運動学習と上肢の運動学習～. 筑波大学附属桐が丘養護学校研究紀要, 第37巻, 49-62.
7. 石川紀宏 (2002) 個別指導計画に関する評価研究2～Evidence Based Educationへの挑戦～. 筑波大学附属桐が丘養護学校研究紀要, 第38巻, 105-114.
8. 石川紀宏 (2003) 個別指導計画に関する評価研究3～自立活動における集団活動の評価研究～. 筑波大学附属桐が丘養護学校研究紀要, 第39巻, 111-122.
9. 石川紀宏 (2005) 自立活動に関する評価研究～ある児童の学習調査に基づく自立活動から教科学習への展開～. 筑波大学附属桐が丘養護学校研究紀要, 第41巻, 7-16.
10. 石川紀宏 (2005) 肢体不自由と身体運動～自立活動における継次処理と同時処理の学習課題に関する探索的研究～. 筑波大学附属桐が丘養護学校研究紀要, 第41巻, 17-24.
11. 石川紀宏 (2007) 子どもの新しい実態把握方法の提起～「KJ法」による実態把握の批判的検討を介して～. 筑波大学附属桐が丘特別支援学校研究紀要, 第43巻, 53-61.
12. 澤田明実・石川紀宏 (2007) 漢字カードを利用した漢字の読字力の形成～視知覚に困難のある脳性まひ児の事例～. 筑波大学附属桐が丘特別支援学校研究紀要, 第43巻, 69-73.
13. 石川紀宏・澤田明実 (2008) 挨拶行動を改善するための一事例計画. 筑波大学附属桐が丘特別支援学校研究紀要, 第44巻, 64-72.
14. 石川紀宏 (2009) 肢体不自由児における認知型の再検討. 筑波大学附属桐が丘特別支援学校研究紀要, 第45巻, 41-46.
15. 全国肢体不自由養護学校校長会 (2005) 『特別支援教育に向けた新たな肢体不自由教育実践講座』. ジェームス教育新社
16. 田中敏・山際勇一郎 (1992) 『新訂ユーザーのための心理・教育統計と実験計画法』. 教育出版
17. 田中敏・中野博幸 (2004) 『実践データ解析法 10秒でできるクイックデータアナリシス』. 新曜社
18. G.K. カンジ (池谷祐二・久我菜穂子訳) (2009) 『「逆」引き統計学 実践統計テスト100』. 講談社
19. Conover, W.J (1999) Practical Nonparametric Statistics, 3rd ed. John Wiley & Sons.
20. Siegel, S. and Castellan, John. N. (1988) Nonparametric Statistics for the Behavioral Sciences, 2nd ed. McGraw-Hill.