

統計学入門講義の再履修者における 統計学の基礎知識についての調査

A Survey of Students Retaking an Introductory Statistics Course on the
Basic Knowledge of Statistics

佐藤柚弥^{†1} 寺尾敦^{†2}

Yuya Sato, Atsushi Terao

^{†1} 青山学院大学大学院社会情報学研究科

Graduate School of Social Informatics, Aoyama Gakuin University

^{†2} 青山学院大学社会情報学部

School of Social Informatics, Aoyama Gakuin University

Abstract: In the School of Social Informatics at Aoyama Gakuin University, freshman students are required to take a course in introductory statistics. Every year, approximately 20 to 30 percent of students fail to pass the course and are therefore required to take it again the following year. The purpose of this study is to assess these students' basic knowledge of statistics. Of the 67 students who retook the course during academic year 2019, 39 students were administered a short test to assess their knowledge at the beginning of the course, and they agreed to participate in this study. The test consisted of several statistical problems adopted from the high school mathematics curriculum. For evaluating their test scores, scores on the same test acquired from 173 freshman students during academic year 2018 were used as a benchmark. Overall, the performance of the students retaking the introductory statistics course was inferior to that of the freshman students. Students who took the same test twice, during both 2018 and 2019, showed some improvement, albeit minimal. These results would appear to indicate that careful instruction is required for students retaking this course in light of their lack of basic knowledge of statistics.

キーワード：統計学，高校数学，質保証，診断的評価，大学機関調査（IR）

Keywords: statistics, high school mathematics, quality assurance, diagnostic evaluation, institutional research

1. 問題と目的

青山学院大学社会情報学部では1年生の必修科目のひとつに「統計入門」がある。記述統計学の基礎から始め、基本的な推定と検定の理論を理解して実

際に使えるようになるまで学習を行う。テキストとしてホーエル（1981）を用いており、第9章までが学習の範囲である。授業は半期15週にわたり、1週あたりの授業時間は180分である。シラバスはウェブ公開されている（<http://syllabus.aoyama.ac.jp/>）。1年生は3クラスに分けられている。そのほかに再履修者のクラスが1クラスある。それぞれ

のクラスは異なった教員が担当し、教え方も異なっているが、獲得しなければならない知識は統一されている。成績評価は2回のテストに基づいて行われ、テストと評価方法はすべてのクラスで同一である。

統計入門の初回授業において、高校数学のカリキュラムで学習する統計学の知識を評価するための小テストを行っている。統計入門の最初に学習する記述統計学の基礎について、現在の中学校および高等学校のカリキュラムはかなりの部分をカバーしている。高等学校までに学習する記述統計学の知識について診断的評価を行うことで、統計入門の学習をどこから始めればよいのか、記述統計学についての学習をどの程度行えばよいのかを考えることができる。平成21年(2009年)に告示された高等学校の学習指導要領に基づき、平成24年度(2012年度)から実施されている高校数学のカリキュラムでは、必修科目である「数学Ⅰ」に「データの分析」という単元がある。「統計の基本的な考えを理解するとともに、それを用いてデータを整理・分析し傾向を把握できるようにする」ことがこの単元の目標である。具体的には、分散、標準偏差、四分位偏差を用いて「データの散らばり」を把握すること、散布図や相関係数を用いて「データの相関」を把握することを学習する。中学校で学習する代表値とあわせて、記述統計学の基礎知識が獲得されることになる。この流れは平成30年告示の学習指導要領に継続されている。

本研究の目的は、記述統計学の基礎についての、統計入門の再履修者の知識を評価することである。統計入門の初回授業で実施している小テストは再履修者も受験している。再履修者を対象にしたクラスでは、初めてこの授業を受ける1年生と同じように教えればよいのか、それとも教え方を変えた方がよいのか、テスト成績をもとに判断を行いたい。再履修者なのだから、昨年度の履修において単位を取得できる水準に達していなかったことは明らかだが、初めて授業を受ける平均的な1年生よりは多くの知識を獲得できているのだろうか？あるいは、平均的な1年生よりも知識が乏しく、1年生に教える以上にいいないな支援が必要なのだろうか？これは診断的評価を行ってみたいとわからない。

1年生の必修科目において学生の知識を診断的に評価することはIR (institutional research) と

しての側面を持つ。IRとは「機関の計画立案、政策形成、意思決定を支援するための情報を提供する目的で、高等教育機関の内部で行われるリサーチ」(Sauepe, 1990)である(この日本語訳は、大塚・山田, 2012)。IRの目的には機関外部への説明責任と機関内部の教育改革・改善がある(大塚・山田, 2012)。外部への説明責任として、入学時にどのような知識を持っていた学生が、どのような知識を獲得して学士課程を修了したのかを明示することは重要である。大学教育の質保証のため、IRにおいて学習成果測定は強い関心を集めている。調査やテストを用いて学習前の状況を明らかにしておくことで、学習成果をより明確に説明できる。内部の教育改革・改善として、1年生の必修科目の単位を取得できなかった学生の知識を測定することは、再履修者クラスでの授業をデザインするための重要な情報を提供する。

2. 方法

2.1 調査協力者

本論文で分析した「統計入門」再履修者のデータは、2019年9月24日の統計入門の初回授業で行われた調査への解答である。調査は小テストの形式で行われた。調査は授業の一部として実施されたので、出席していたすべての学生(1年生および再履修学生)が解答した。小テストに解答した再履修学生は52名であり、このうちデータおよび分析結果を公表することに同意した39名(性別:男性28名、女性11名;学年:2年生31名、3年生8名;年齢:平均20.0歳、標準偏差0.92歳)から得られたデータを分析した。統計入門に受講登録されている再履修者は、2年生50名、3年生17名の、67名であった。よって、再履修者全体に占める調査協力者の割合は58%であった。

参加同意書に記入された生年月日から判断すると、39名の再履修者はすべて、高等学校の数学Ⅰ「データの分析」を学習していたはずである。よって、本調査で使用された問題に正解するために必要な知識は、統計入門のみならず、高等学校の数学Ⅰでも学習していたはずである。

再履修者の小テスト成績を評価するため、2018年9月18年の統計入門の初回授業で行われた同一の小テストに解答し、データおよび分析結果を公表することに同意した1年生（当時）173名の解答を使用した。この年度の1年生の成績を用いたのは、再履修者のうちの多くを占める2年生と入学年度（2018年度）が同じ学生との比較で履修者の成績を評価するためと、個人内での小テスト成績の変化を検討するためである。再履修者である2019年度の2年生31名のうち、29名は1年生であった2018年度にも小テストに解答し、データおよび分析結果を公表することに同意していた。この29名については1年次と2年次の成績を比較することが可能である。

2.2 材料

1年生に実施するために、調査（小テスト）は高等学校の数学Iまでに学習した知識で正解できる問題から構成された。2018年度および2019年度ともに同じ問題を使用した。使用した問題を付録1に示す。

問題1は、平均値、中央値、最頻値という3つの代表値の区別に関する問題であった。この問題は、2011年4月から7月にかけて日本数学会が大学生を対象に行った「大学生数学基本調査」で使用された3問のうちのひとつ（問1）を改変したものであった。改変は3か所を行った。第1に、選択肢と解答方法を変更した。もとの出題では[1]から[3]の3つの叙述があり、解答方法が「確実に正しいと言えることには○を、そうでないものには×を、左側の空欄に記入してください」と指示されていた。今回の小テストでは、[1]から[3]のいずれも正しくないという叙述を[4]として加え、問題文での解答方法の指示を「以下の選択肢[1]から[3]のうち、この結果から確実に正しいと言えることだけを選んでください（複数選択可）。選択肢[1]から[3]に正しいものがひとつもないときには、選択肢[4]を選んでください」に変更した。第2に、提示されている163.5 cmという平均値が、四捨五入のない正確な数値であることを明示した。もとの調査では「その平均を計算すると163.5cmになりました」であった記述を、「その平均を計算すると、ちょうど163.5cmになりました（四捨五入なし）」に変更した。これは[2]を確実に正し

い記述とするための変更であった（たとえば、100人の身長合計が16346cmで、小数点第2位を四捨五入して平均値を163.5cmとしたという可能性を排除している）。第3の変更として、[1]から[3]の選択肢を正しいあるいは誤っていると判断した理由を記述することを求めた。

問題2は、小さなデータセットが与えられ、要約統計量（平均値、中央値、最頻値、分散、標準偏差）を計算する問題であった。これは、受験参考書である『30日完成！センター試験対策 数学IA』（チャート研究所、2014、p.70）に収録されていた問題の一部である。もとの問題では数値のみ解答することが求められていたが（センター試験対策の受験参考書であるため、マークシートへの解答が想定されていた）、途中の計算や解答の根拠を記述するように改変した。

問題3は与えられたいくつかの統計量から適切なものを選択して相関係数の計算式を示す問題であった。これは2015年1月に実施されたセンター試験本試験での数学IA第3問の一部である。もとの問題の一部のみ使用したためと、解答をマーク式でなく記述としたため、問題として成立するよう改変した。

小テストの問題は1冊の冊子に印刷された。表紙に続き、それぞれの問題は見開きの2ページを用いて印刷された（問題3は左側のページのみ使用した）。裏表紙にはこの調査への参加の同意書が印刷された。

2.3 手続き

統計入門の初回授業では、1年生および再履修者の全員がひとつの教室に集められ、ガイダンスが行われた。

ガイダンスに続いて調査（小テスト）が実施された。調査の目的は、統計入門の開始時点での統計学に関する履修者の知識を調査し、今後の授業を進めるうえでの参考とすることであると説明がされた。小テストの問題冊子が教員およびアシスタントによって配布された。配布された冊子は開かず、表紙に印刷された注意事項を読むように指示がなされた。

受講者は全員が同じペースで3つの問題それぞれに解答した。すなわち、解答開始の指示で全員が解答を始め、終了の指示で解答を終えた。3つの問題の解答時間はそれぞれ10分、10分、5分であった。与

えられた制限時間よりも前に解答を終えたときには、次の問題に進まずそのまま待つように指示がされた。

第3問への解答を終えた後、データおよび分析結果を公表することに同意できる場合には、裏表紙にある同意書に署名をすることを求めた。

3. 結果と考察

3.1 スコアリング

問題1から問題3で構成された小テストには、正誤を容易に判定できる小問が7つ含まれていた。これらそれぞれについて正解に1点を与えた。よって、小テストのスコアは7点満点となった。

問題1では、正しい記述の選択肢として[2]だけを選んでいれば1点を与えた。これ以外の選択肢を選んだ場合や、選択肢[2]と共にほかの選択肢も選んだ場合は不正解であり、0点とした。

問題2では、5つの統計量それぞれに対して、正しい値が示されている場合に1点を与えた。問題2で問われている統計量は5つあったので、これらすべての統計量を正しく解答すれば5点となった。正しい分散の値が解答されていない場合は、標準偏差も誤りとした。ただし、分散の値が誤っているときに、その正の平方根を標準偏差として解答した場合（たとえば、偏差平方和である40を分散とした誤答で、標準偏差を $\sqrt{40}$ と解答した場合）は、白紙などその他の誤答と区別してデータの記録を行った。

問題3では、相関係数を計算する正しい計算式、すなわち、 $54.30/(8.21 \times 6.98)$ という解答に1点が与えられた。出題では2変数の標準偏差だけでなく分散の値も示されていたので、 $54.30/(\sqrt{67.40} \times \sqrt{48.72})$ という解答も正解となる。再履修者ではこの解答を与えた学生はいなかった。

問題1では、3つの選択肢それぞれについて、正しい、あるいは、誤り（必ずしも正しくない）と判断した理由の記述を求めている。2人の評定者（本論文の2名の著者）が、再履修者39名による記述を3つのカテゴリに分類した。分類された記述の総数は117（ $=39 \times 3$ ）であり、2人の評定者の分類は3つの記

述を除いて一致した。判断が一致しなかった記述は2人が議論してカテゴリを決定した。カテゴリ分類は2段階で行われた。2人の評定者は、最初に、明らかな誤答を「誤答」カテゴリに分類した。選択肢[1]から[3]のいずれかで正誤判断を誤っている場合は、当然ながらその判断理由の記述も誤ったものになった。次に、選択肢[1]と[3]については、「誤答」ではない解答をさらに2つに分類した。選択肢[1]は、中央値については正しい（提示されている163.5cmという数値が中央値であるならば正しい）が、平均値については必ずしも正しくない。そこで、平均値は中央値と異なるという内容の記述を「正解A」とし、それ以外の解答を「正解B」とした。選択肢[3]は、最頻値については正しいが、平均値については必ずしも正しくない。平均値是最頻値と異なるという内容の記述を「正解A」とし、それ以外の解答を「正解B」とした。「正解B」カテゴリに分類された解答には、明らかな誤答ではないが、選択肢の正誤を判断した根拠としては不十分なもの（たとえば、「問題文の記述だけでは正しいとは言えない」といった、判断根拠の記述になっていないもの）も含まれている。選択肢[2]では、明らかな誤答ではない解答は、「中学校の3年生の生徒100人の身長平均は身長の合計を100で割ったものだから正しい」といった、平均値の定義式に言及した1種類しか実質的に見られなかった。この解答を「正解A」とした。

再履修者39名の解答をスコアリングして得られたデータを付録2に示す。

3.2 全体的な成績

2019年度再履修者と2018年度1年生の、各問題の正解者数と正解率を表1に示す。

代表値についての問題での再履修者の正解率は、それほど悪くはないが、1年生と比べて相対的に低かった。代表値の区別（問題1）での再履修者の正解率は79%、1年生の正解率は84%であった。代表値の計算（問題2での3つの小問）については、平均値の計算は再履修者全員が正解していた（再履修者100%、1年生97%）。中央値の計算（問題文で与えられた測定値を大きさの順に並べ替えれば得られる）では、再履修者の正解率は77%、1年生の正解率

表 1 各問題の正解者数および正解率

問題	2019 年度再履修者 (N=39)		2018 年度 1 年生 (N=173)	
	正解者 (名)	正解率 (%)	正解者 (名)	正解率 (%)
1. 代表値の区別	31	79	146	84
2. 統計量の計算				
平均値	39	100	168	97
中央値	30	77	145	84
最頻値	34	87	159	92
分散	7	18	63	36
標準偏差	7	18	55	32
3. 相関係数の計算	5	12	26	15

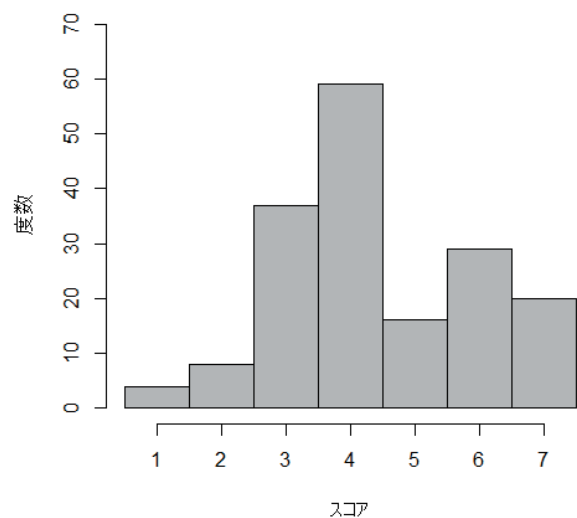
Note. 2019 年度の再履修者 39 名のうち、29 名は 2018 年度に 1 年生として小テストに解答していた。

は 84% であった。最頻値については、再履修者の正解率は 87%、1 年生の正解率は 92% であった。

散布度および相関係数については、再履修者の知識に大きな問題が認められた。統計量の計算（問題 2）のうち分散と標準偏差、および、相関係数の計算（問題 3）は、2019 年度再履修者と 2018 年度 1 年生ともに、60% 以上の学生が不正解であった。分散および標準偏差の計算では、1 年生の正解率は分散 36%、標準偏差 32% でかなり低いが、再履修者はさらに低く、分散 18%、標準偏差 18% であった。相関係数の計算の正解率は再履修者も 1 年生もかなり低く、再履修者は 12%、1 年生は 15% であった。分散を正しく求めることのできていた再履修者 7 名は標準偏差の計算式も正しく記述できていた。分散の計算は誤っていたが（白紙の誤答も含む）、標準偏差が分散の正の平方根となることを記述していた再履修者が 8 名いた。このことは知っていても、分散の計算ができなかったために標準偏差の計算が白紙回答となった学生もいるだろう。よって、再履修者の多くは標準偏差が分散の正の平方根であることは理解しているかもしれない。

小テストのスコアの分布を検討すると、再履修者では比較的高いスコアの学生が少なかった。2019 年度再履修者と 2018 年度 1 年生それぞれについて、小テストのスコアのヒストグラムを図 1 に、スコアの要約統計量を表 2 に示す。スコアのとりうる値は 0 から 7 までの整数値である。ヒストグラムを見ると、1 年生に比べ、再履修者では最頻値（4 点）を超える 5 点以上の学生が少ないことがわかる。実際、5 点以上

2018年度1年生



2019年度再履修者

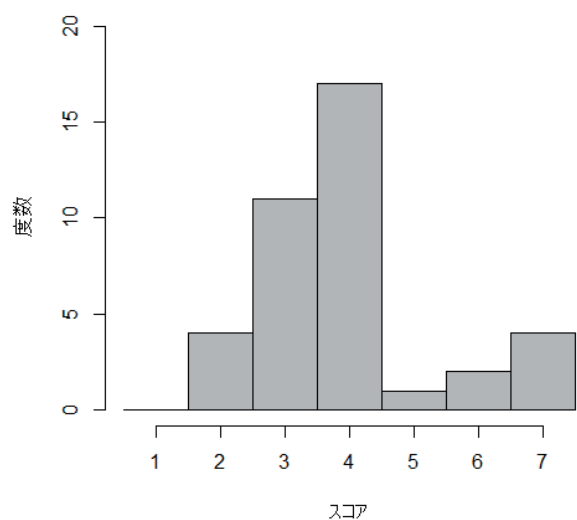


図 1 テストスコアの分布

表 2 テストスコアの要約統計量

要約統計量	2019 年度再履修者 (N=39)	2018 年度 1 年生 (N=173)
平均値	3.9	4.4
中央値	4.0	4.0
最頻値	4	4
最小値	2	1
最大値	7	7
標準偏差	1.4	1.5
四分位偏差	0.5	1.5

Note. テストスコアは 7 点満点の整数値をとる。
標準偏差の計算では偏差平方和を測定値の数 (N) で割った。

表 3 代表値の性質の正誤判断における根拠の記述の分布

選択肢	記述カテゴリ		
	正解 A	正解 B	誤答
[1] 中央値	9 (23%)	20 (51%)	10 (26%)
[2] 平均値	32 (82%)	-----	7 (18%)
[3] 最頻値	3 (8%)	23 (59%)	13 (33%)

Note. 平均値については、正解のカテゴリは 1 種類（正解 A）のみである。

の学生の割合は 1 年生で 38% (65 名) であったが、再履修者では 18% (7 名) であった。表 1 からわかるように、再履修者では、統計量の計算（問題 2）のうち分散と標準偏差、および、相関係数の計算（問題 3）という、3 つの問題での正解率が低い。このことが 5 点以上の学生の割合を引き下げたと考えられる。再履修者と 1 年生ともにスコアの中央値と最頻値は 4 点だが、平均値は再履修者の方が 0.5 点低かった（再履修者 3.9 点、1 年生 4.4 点）。この結果も、分散、標準偏差、相関係数を計算できない再履修者が相対的に多いことを反映していると考えられる。

3.3 代表値の性質についての理解

問題 1 で問われている代表値の性質について、再履修者の理解をより詳細に検討した。この問題では、代表値の性質についての選択肢 [1] から [3] のうち確実に正しいものを選択するとともに、これら選択肢を正しい、あるいは、誤っている（必ずしも正しくない）と判断した理由の記述を求めている。再履修者によるこの記述を 2 人の評定者が分類した結果を表 3 に示す。方法のセクションで述べたように、明らかな誤り

のある記述を「誤答」と分類し、その他の記述を「正解 A」と「正解 B」に分類した。正解 A は、選択肢 [1] および [3] についてはそれぞれ、中央値および最頻値に言及した記述である。選択肢 [2] では測定値の合計を標本の大きさを割ったものが平均値であるということに言及した記述である。誤答でなく、正解 A でもない記述を正解 B とした。

選択肢 [1] の判断理由では 10 名 (26%)、[2] では 7 名 (18%)、[3] では 13 名 (33%) が誤答であった。記述 [2] のみを正しいと解答していても、[1] から [3] での理由の記述に明らかな誤りを含む学生を不正解として扱えば、問題 1 の正解者は 21 名 (54%) となった。理由の評定では明らかな誤りだけを「誤答」カテゴリに分類したので、もっと厳しく評定すれば正解率はさらに低下するだろう。問題 1 で問われている代表値の性質を正確に理解できている再履修者はせいぜい半数程度であると考えられる。

選択肢 [1] および [3] では、中央値あるいは最頻値に言及した正解 A よりも、その他の正解 B の方が多かった。平均値と中央値の違い、および、平均値と最頻値の違いについては、高校の数学 I で学習しているし、1 年次の統計入門でも学習している。しかし、

この違いについて明確に理解できている再履修者は相対的に少ないのかもしれない。

正しい選択肢 [2] の判断理由はかなり正解率が高く、82%であった。それでも、この選択肢は平均値の定義式から直ちに正しいとわかるので、もっと高い正解率であってもよさそうに思われる。選択肢 [2] を誤りであると考えた再履修者には、この選択肢が正しいとすると全員の身長が平均値と同じということになると記述した。かなり深刻な誤解をしている学生が4名いた。

3.4 講義履修の効果

2019年度の再履修学生において、1年次に統計入門の講義を受けたことの効果が認められるかどうかを検討した。2019年度の2年生31名のうち、29名

は1年生であった2018年度にも小テストに解答し、データおよび分析結果を公表することに同意していた。そこで、この29名について、1年生のときのスコアと2年生のときのスコアを比較した。これら最履修者のテストスコアの分布を図2に示す。上段の2つのパネルは年次別に描いたヒストグラムである。下段は、29名の学生それぞれについて2年次（2019年度）のスコアから1年次（2018年度）のスコアを引いて得られた、変化量のヒストグラムである。年次別のテストスコア、および、テストスコアの変化量の要約統計量を表4に示す。

2つの年度のスコアについて、対応のある2標本での母集団平均の差の検定（有意水準5%、両側検定）を行うと、 $t(28)=1.887$, $p=.070$, $d=0.35$, 95% CI [-0.33, 0.79] となった。統計入門の学習前後での比較なので片側検定でもよさそうに思えるが、

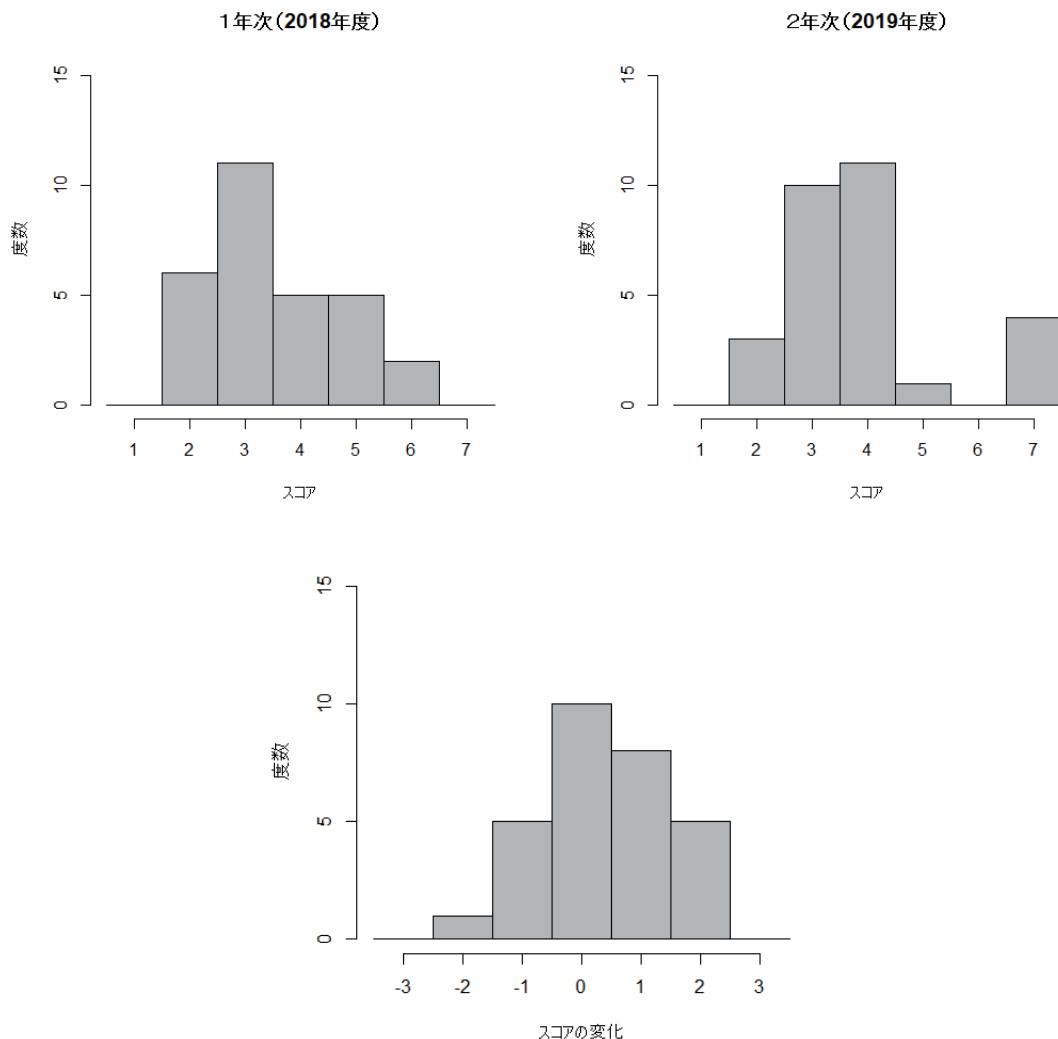


図2 再履修者の年次別テストスコアおよびスコアの変化の分布 (N=29)

表 4 2 年生再履修者の年次別テストスコアおよびスコアの変化の要約統計量

要約統計量	年次		変化
	1 年(2018)	2 年(2019)	
平均値	3.5	3.9	0.4
中央値	3.0	4.0	0.0
最頻値	3	4	0
最小値	2	2	-2
最大値	6	7	2
標準偏差	1.2	1.4	1.1
四分位偏差	0.5	0.5	1.0

Note. 標準偏差の計算では偏差平方和を測定値の数 ($N = 29$) で割った。
スコアの変化は学生それぞれについて計算し、要約統計量を求めた。

時間の経過とともに知識が非活性となる可能性も考えて両側検定とした。効果量 d は、学生ごとに計算されたスコアの変化量について、平均値 (0.3793) を標準偏差 (1.0828) で割った値である。この標準偏差は不偏分散の正の平方根である。統計的仮説検定の結果は有意ではなく、変化量の母集団平均が 0 であるという帰無仮説は棄却できなかった。ここでの母集団としては、現在の入学者層が変化せず、統計入門の学習内容も変化しない場合に、統計入門を再履修することになる学生全体を想定している（つまり、2019 年度の再履修者に限定されず、今後の潜在的な再履修者を含む）。図 2 および表 4 を見ると、2 年次 (2019 年度) の成績は 1 年次 (2018 年度) よりも向上している印象を受けるかもしれない。代表値 (平均値、中央値、最頻値) はいずれも 2 年次の方が高い。1 年次は満点者がいなかったが 2 年次には 4 人いる。統計的仮説検定の結果は有意水準 10% ならば有意である（これは「有意傾向」と表記されることがある）。しかし、効果量はそれほど小さくなく、母集団での変化量の平均は高く見積もって 1 点に満たない (95% 信頼区間の上側信頼限界が 0.79 である)。したがって、1 年次に統計入門の講義を受けたことの効果はわずかに認められるものの、非常に小さいと言える。

ていた。代表値についての問題での再履修者の正解率は、それほど悪くはなかったが、1 年生と比べて相対的に低かった。散布度および相関係数については、再履修者の知識に大きな問題が認められた。分散と標準偏差の計算は、1 年生での正解率も低かったが、再履修者の成績はさらに低かった。相関係数の計算の正解率は 1 年生も再履修者もかなり低かった。

代表値の性質について再履修者の理解を詳細に検討すると、3 つの代表値 (平均値、中央値、最頻値) の性質を正確に理解し、これらを区別できる再履修者は、せいぜい半数程度であった。

再履修者の 1 年次のテストスコアと 2 年次のテストスコアを比較すると、改善はごくわずかで、統計入門を履修したことの効果はほとんど認められなかった。

こうした結果は、再履修者のクラスでの統計入門の授業は、1 年生に対する授業と同じかそれ以上にいていねいな学習支援が必要であることを意味する。再履修者の多くは、高等学校で学習し、1 年次の統計入門でも学習した記述統計学の知識を十分に獲得していない。再履修者クラスの授業では、これまでに統計学を学習してきたことを前提とせず、記述統計学の基礎から教示を行う必要があるだろう。

謝辞

本論文は JSPS 科研費 17K04356 の助成を受けた研究「R を用いた心理統計教育に活かす Web 教材作成とその効果の検証」(代表者: 山田剛史) の一部である。

4. 結論

統計入門の再履修者が持つ統計学の基礎知識は、この講義を初めて履修する 1 年生よりも全体的に劣っ

引用文献

チャート研究所(編著)(2014). 30日完成! センター試験対策 数学IA 数研出版

ホーエル, P. G. (1981). 初等統計学 培風館

日本数学会教育委員会 (2013). 第一回 大学生数学基本調査報告書

<https://mathsoc.jp/publication/tushin/1801/chousa-houkoku.pdf>

大塚雄作・山田剛史 (2012). 大学教育評価 京都大学高等教育研究開発推進センター(編) 生成する大学教育学 第5章 (pp. 165-216) ナカニシヤ出版

Saupe, J. L. (1990). The functions of institutional research (2nd ed.). Association for Institutional Research. Washington D. C.: Microfiche.

付録 1 調査に用いた問題

調査に用いた3つの問題を以下に示す。3つの問題が1冊の冊子に印刷された。表紙に続き、それぞれの問題は見開きの2ページを用いて印刷された(問題3は左側のページのみ使用した)。実際の問題冊子では小問ごとに解答欄が設けられていたが、以下では省略した。裏表紙にはこの調査への参加の同意書が印刷された。

問題 1:

ある中学校の3年生の生徒100人の身長を測り、その平均を計算すると、ちょうど163.5cmになりました(四捨五入なし)。

(1) 以下の選択肢[1]から[3]のうち、この結果から確実に正しいと言えることだけを選んでください(複数選択可)。選択肢[1]から[3]に正しいものがひとつもないときには、選択肢[4]を選んでください。

[1] 身長が163.5cmよりも高い生徒と低い生徒は、それぞれ50人ずついる。

[2] 100人全員の身長をたすと、 $163.5\text{cm} \times 100 = 16350\text{cm}$ になる。

[3] 身長を10cmごとに「130cm以上で140cm未満の生徒」「140cm以上で150cm未満の生徒」・・・というようにに区分けすると、「160cm以上で170cm未満の生徒」が最も多い。

[4] 選択肢[1]から[3]に正しいものはない。

(2) 選択肢[1]について、あなたが「正しい」あるいは「誤っている」と考えた理由を説明してください。(選択肢[1] 身長が163.5cmよりも高い生徒と低い生徒は、それぞれ50人ずついる。)

(3) 選択肢[2]について、あなたが「正しい」あるいは「誤っている」と考えた理由を説明してください。(選択制[2] 100人全員の身長をたすと、 $163.5\text{cm} \times 100 = 16350\text{cm}$ になる。)

(4) 選択肢[3]について、あなたが「正しい」あるいは「誤っている」と考えた理由を説明してください。(選択肢[3] 身長を10cmごとに「130cm以上で140cm未満の生徒」「140cm以上で150cm未満の生徒」・・・というようにに区分けすると、「160cm以上で170cm未満の生徒」が最も多い。)

問題 2:

(1) 次のデータは、あるソフトボールチームの15試合の得点である。

7, 3, 0, 1, 4, 1, 7, 6, 5, 6, 6, 2, 4, 6, 5

このデータの平均値、中央値、最頻値を求めよ。数値だけを解答するのではなく、途中の計算や、解答の根拠を以下の枠内に必ず記述すること。

(2) あるクラスの生徒 5 人に、数学の小テストを実施したところ、次のような結果となった。

5, 8, 10, 9, 8 (単位は「点」)

この得点の分散と標準偏差を求めよ。

分散の計算では、数値だけを解答するのではなく、途中の計算や、解答の根拠を以下の枠内に必ず記述すること。

標準偏差については、分散の値を使った計算式だけを示せばよい。

問題 3:

ある高校 2 年生 40 人のクラスで、1 人 2 回ずつハンドボール投げの飛距離のデータをとった。以下に示された数値を用いて、相関係数を計算するための計算式を書け。計算を実行して相関係数を求める必要はない。

	平均値	中央値	分散	標準偏差
1 回目のデータ	24.70	24.30	67.40	8.21
2 回目のデータ	26.90	26.40	48.72	6.98

1 回目のデータと 2 回目のデータの共分散	54.30
------------------------	-------

付録2 再履修者のデータ

SID	Q1	R1	R2	R3	Q2m	Q2md	Q2mo	Q2v	Q2sd	Q3	S18
1	1	B	A	B	1	1	1	0	0	0	NA
2	1	W	A	W	1	1	1	1	1	0	NA
3	1	W	A	B	1	1	1	0	0	0	NA
4	1	W	A	B	1	1	1	0	0.5	0	NA
5	1	B	A	B	1	1	1	0	0	0	NA
6	1	B	A	B	1	1	1	1	1	0	NA
7	1	A	A	W	1	1	1	0	0	0	NA
8	0	A	W	W	1	1	0	0	0	0	NA
9	1	B	A	B	1	1	1	0	0	0	NA
10	0	B	W	B	1	1	1	0	0.5	0	4
11	1	B	A	A	1	1	1	0	0	0	5
12	1	B	A	W	1	1	1	0	0.5	0	3
13	0	W	W	W	1	0	1	0	0	0	2
14	1	B	A	B	1	0	1	0	0	0	NA
15	1	B	A	B	1	0	0	1	1	0	2
16	1	A	A	B	1	1	1	1	1	1	5
17	1	B	A	B	1	1	1	1	1	1	5
18	0	W	A	W	1	0	1	0	0	0	2
19	1	W	W	W	1	1	0	0	0	0	3
20	1	W	A	W	1	0	1	0	0	0	5
21	0	B	A	W	1	1	1	0	0	0	3
22	1	A	A	A	1	1	1	1	1	1	6
23	0	W	W	B	1	0	1	0	0	0	3
24	1	B	A	B	1	1	1	0	0	0	4
25	0	W	W	B	1	1	1	0	0	0	2
26	1	B	A	B	1	1	1	0	0.5	0	3
27	1	B	A	B	1	1	1	0	0.5	1	3
28	1	W	A	W	1	1	1	1	1	1	6
29	0	A	W	W	1	1	1	0	0	0	2
30	1	B	A	B	1	1	1	0	0	0	5
31	1	A	A	B	1	1	1	0	0.5	0	3
32	1	B	A	B	1	0	1	0	0	0	3
33	1	A	A	W	1	1	1	0	0.5	0	2
34	1	B	A	B	1	0	1	0	0	0	3
35	1	A	A	B	1	1	1	0	0	0	3
36	1	B	A	W	1	1	1	0	0.5	0	4
37	1	A	A	A	1	1	1	0	0	0	4
38	1	B	A	B	1	0	1	0	0	0	3
39	1	B	A	B	1	1	0	0	0	0	4

表の最上段は変数ラベルである.

SID: 通し番号

Q1: 問題 1 の正誤. 1 が正答, 0 が誤答を意味する.

R1 から R3: 問題 1 の選択肢 [1] から [3] の正誤判断における根拠の記述の分類. A は正解 A, B は正解 B, W は誤答.

Q2m: 問題 2 (1) での平均値の正誤. 1 が正答, 0 が誤答を意味する.

Q2md: 問題 2 (1) での中央値の正誤. 1 が正答, 0 が誤答を意味する.

Q2mo: 問題 2 (1) での最頻値の正誤. 1 が正答, 0 が誤答を意味する.

Q2v: 問題 2 (2) での分散の正誤. 1 が正答, 0 が誤答を意味する.

Q2sd: 問題 2 (2) での標準偏差の正誤. 1 が正答, 0 が誤答を意味する. 0.5 は, 分散の計算は誤り(あるいは無回答)だったが, 標準偏差が分散の正の平方根であるという記述がされていた解答である.

Q3: 問題 3 の正誤. 1 が正答, 0 が誤答を意味する.

S18: 2018 年度にも同じ小テストを受験していた 2 年生 29 名の, 1 年次 (2018 年度) のスコア. 3 年生の再履修者および 2018 年度のテストの未受験者は NA と記録されている.