

# 統計的検定学習支援用ウェブアプリケーション BuMocの開発

朴 壽永<sup>1,a)</sup> 新部 昭夫<sup>2</sup> 安江 紘幸<sup>3</sup> 井形 雅代<sup>2</sup> 山田 崇裕<sup>2</sup>

受付日 2019年9月9日, 採録日 2020年2月4日

**概要:** 2008年に日本においても新しい学習指導要領が改訂され, 公式を暗記して計算練習する旧来の統計教育ではなく, データに基づく科学的な問題解決力をコンピテンシーとして定着させる, 新しい枠組みの下での統計教育を指向している. しかし, 一般に, データ分析の経験がなく, 抽象化された演繹的思考法を主に教授してきた学校の教員にとって, どのようにこの授業を展開していけばいいのか, 現場の戸惑いの声は大きい. そこで, 統計分析の知識や経験がなくても, 計算の仕方や数式, 公式が分からなくても, 適切かつ手順をふんだ関連のある複数の検定結果が同時に得られ, また, 採用できる検定結果とできない検定結果が視覚的に見分けられる機能を備えた統計的検定の学習支援システムを開発した. 本システムを用いることで, 比較したい群をドラッグアンドドロップで指定するだけで正規性の検定から多重比較まで23種類の検定結果が得られる. 受講者数100人を超える大学3年生の授業でシステムの有用性を確認した. 授業実施前と後における対応のある差の検定結果, 高い有意差が認められた. ユーザはシステムのマニュアルを読むことで自主的に各種検定手法を学ぶことができることが示唆された.

**キーワード:** データサイエンス, 統計思考力, 統計リテラシー, 問題解決能力, リサーチリテラシー

## Development of BuMoc: a Web Application for Learning support of Statistical Tests

SOO-YOUNG PARK<sup>1,a)</sup> AKIO NIBE<sup>2</sup> HIROYUKI YASUE<sup>3</sup> MASAYO IGATA<sup>2</sup> TAKAHIRO YAMADA<sup>2</sup>

Received: September 9, 2019, Accepted: February 4, 2020

**Abstract:** In 2008, the new course of study was revised in Japan, instead of traditional statistical education to practice calculation by memorizing the formula, it is oriented towards statistical education under a new framework of establishing scientific problem solving ability based on data as a competency. In general, however, the school teachers have mainly taught the abstracted deductive thinking method without data analysis experience. There is also great apprehension from them because they do not know how to develop this lesson. Therefore, we developed a web application for learning support of statistical tests. Even if they do not have knowledge or experience in statistical analysis, they can obtain multiple relevant test results that are appropriate and procedural at the same time. In addition, the web application has a function that can visually distinguish between test results that can be adopted and test results that can not be taken. Using the web application, 23 types of test results are obtained from the test of normality to multiple comparison only by specifying the group to be compared by drag and drop. We confirmed the usefulness of the web application among 100 third year university student participants. As a result of statistical significance test of the difference between responses before and after the class, a very high significant difference was recognized. Furthermore, it is suggested that users can learn various test methods voluntarily by reading the manual of the web application.

**Keywords:** data science, statistical thinking, statistical literacy, problem solving ability, research literacy

<sup>1</sup> 県立広島大学  
Prefectural University of Hiroshima, Hiroshima 734-8558,  
Japan

<sup>2</sup> 東京農業大学  
Tokyo University of Agriculture, Setagaya, Tokyo 156-8502,  
Japan

<sup>3</sup> 農研機構東北農業研究センター  
NARO Tohoku Agricultural Research Center, Morioka,  
Iwate 020-0198, Japan

a) park@pu-hiroshima.ac.jp

## 1. はじめに

2007年に、「21世紀の知識創造社会に向けた統計教育推進への要望」[1]が提言され、学習指導要領の数学科の中で「資料の活用」と「データ分析」の単元が設定され、中・高等学校で数学の1つの項目として、統計思考力の育成教育が始まった[2]。垣花[2]によると、統計思考力の育成は統計リテラシーの教育であり、このような不確実性の事象の扱いを含め、基礎的な統計知識をもとに、情報をまとめ、分析し、バラツキや確率を考慮し、問題解決をしたり、論理的に説明したり、批判的考えができる力を育成することになる。

他方、社会科学において、答えが未知である問いを解決しようとする研究の基礎力であるリサーチリテラシーを教える機会が重要であるとされており[3]、谷岡[4]は、社会調査におけるリサーチリテラシーのことを「事実や数字を正しく読むための能力」と呼んでいる。また片瀬[5]は、社会調査教育において分析技法の教育にさらに力を注ぐ必要があるとしながら、リサーチリテラシーに統計リテラシーを合わせた形で統計的リサーチリテラシーという言葉を用いている。

以上より統計リテラシーにおける統計とはリサーチリテラシーに含まれる機能の1つの分析技法であることから、アンケート調査機能と統計分析機能を備えたツールの開発を目的とする本稿においても、片瀬[5]にならい、統計リテラシーとリサーチリテラシーの両方の意味を持つ統計的リサーチリテラシーを用語として用いた。

### 1.1 統計的リサーチリテラシー教育の現状と課題

アメリカ数学会のカリキュラムアクションプロジェクトが発表した1992年のCobbレポート[6]は、推奨事項として、①活用を意識した統計思考力(統計的問題解決力)を重視する、②理論や公式、計算方法と計算演習は控えめにし、データそのものや統計的な考え方を重視する、③問題解決の活動学習を行う、の3点をまとめている[7]。また、1996年に、米国統計学会と全米数学協議会の共同カリキュラム委員会が発した共同指針の中には、「現実似せたデータではなく、実際の生データを使う」、「計算の仕方を教えるよりもコンピュータを使う」、「数式、公式の導出はあまり重要ではない」とされている[8]。

渡辺[7]によると、これまでの統計教育は、すでに訓練された研究者や技術者、もしくは研究を始める大学生などに、統計の用語と統計的方法の適用の仕方(計算方法と手順)を教えることが主であった。しかし、欧米における新しい枠組みでの統計教育は、統計思考力を科学技術推進の第3の腕(the third arm: なくても生きるうえで困らないが、あると飛躍的に効率が上がる手段)として位置付けられており、学校教育の早期より、児童・生徒に、身の回りの問題の解決に統計を結び付けて思考させ、解釈させ、さらに

新しい仮説の創造に至る大きな流れを、繰り返し経験させることで思考力を育成する方式を取っている。一般に、不確実性をともなう現実の問題をデータのバラツキ(分布)で表現し確率モデルに基づいて推論(予測)する統計的な概念は、大学生や大人にとっても理論だけで学ぶことは困難である。したがって、欧米では、学校教育の早い段階から、身の回りのデータに親しみ、目的に沿ったデータの収集と分析、結果の解釈・議論という一連のプロセスを繰り返し経験させ、日常のアクションに統計リテラシーの素養を結び付ける教育カリキュラムとなっている。

一方、東証一部、二部の上場全企業を対象に国立教育政策研究所が行った「企業の算数・数学教育への期待」調査報告[9]では、回答企業の9割以上が学校教育に期待する算数・数学内容として、「データに基づいて予測すること」を「特に大切」、「特に大切にしている部課・部署がある」と回答している。

日本においても、これらに後押しされる形で、文部科学省中央教育審議会[1]の提言に基づき、2008年に新しい学習指導要領[10]が改訂され、公式を暗記して計算練習する旧来の統計教育ではなく、データに基づく科学的な問題解決力をコンピテンシーとして定着させる、新しい枠組みの下での統計教育を指向している。しかし、一般に、データ分析の経験がなく、抽象化された演繹的思考法を主に教授してきた学校の教員にとって、どのようにこの授業を展開していけばいいのか、現場の戸惑いの声は大きい[7]。

また、片瀬[5]は、実際の社会調査教育の現場では、たとえば社会調査士育成をめぐっても、大学間の格差が存在するだけでなく、量的調査では検定や多変量解析、質的調査ではフィールドノートやドキュメントの分析技法の教授までは十分になされていない問題点があり、今後はデータ収集の方法だけでなく、こうした分析技法の教育にさらに力を注ぐことが、社会調査教育の課題となると指摘している。

こうした中で滋賀大学が平成29年に、横浜市立大学が平成30年にそれぞれ「データサイエンス学部」を開設した。総務省[11]によると、データサイエンスとは、「データに基づく数量的な思考により、課題を解決に導く能力」である。また、2008年に新しい学習指導要領の改訂もあり、統計教育に関する関心は高まっているものの、具体的な指導方法に苦慮しており、実践的授業の推進を図るために、適切な教材の作成や提供が必要となっている、としている。

### 1.2 統計的リサーチリテラシー教育におけるIT活用

渡辺ら[12]によれば、従来の統計学の講義形態は数理的解説が主であり、データの解析ツールとしての実践的な指導がとまわらない点が少なからず問題であった。しかし、近年のインターネットの普及は、マルチメディアや動的グラフ、データ提供サイトの活用を通して、受動学習から能動型・参加型学習へ、また理論学習から効率的な実践学習

への移行を容易にする新媒體の教材開発を可能にしていると、統計書籍専用検索エンジンの開発、エクセルと分析マクロを利用した教育用統計分析プログラムの開発などを試みた。また、渡辺ら [13] は、統計基礎学習のための e-learning 教材「身近な統計 (Web 版)」を研究開発した。さらに、ドイツ (New Statistics: www.neuestatistik.de)、韓国 (韓国放送大学の Web 配信コース)、中国 (Renmin University of China) との国際協力による Web をベースにした配信教材と教材資源の共有化、母国語でのローカライゼーションも行っている。開発コストが高く予算上の制約をとまうからである。

他には、統計の解説およびユーザが Web 上のテキストボックスにデータを入力することで分析結果が得られる「おしゃべりな部屋 (プラネタリウム, 星, 植物, 熱帯魚, 統計学)」というサイトもあり、当サイトのアクセス数は現在 560 万を超えている [14]。

これらのインターネットおよびイントラネット上の Web 教材は、言語の障壁を設けることは学生にとって閾値を上げるだけで効果的ではないとの共通の認識の下、教材はすべて各国の母国語で開発されている [12]。

### 1.3 統計的リサーチリテラシー教育のための IT を活用したシステム開発の必要性

統計的リサーチリテラシーにおいて、たとえば、統計ソフトの SPSS を用いた 2 群間の差の検定を行う際の手順は通常次のようになる。アンケート調査による生データの収集 → 分析手法の選択 → 変数の挿入など詳細を設定・分析実行のボタンクリック → 示された結果を読み取る、である。ここで「アンケート調査による生データの収集」は、Google フォーム [15] を利用することで比較的簡単にアンケートフォームの作成と実施およびデータ収集ができる。しかし、Google フォームには統計分析機能がないため、改めて SPSS などの統計分析専用ソフトを活用することになるが、上述の「分析手法の選択」の段階において複数の分析ツールの中で自ら適切かつ順序良く手法を適用、実行しなければならない。この段階で、たとえば、データが正規分布に従わないときに、パラメトリック検定を使用してしまう統計学的誤りは研究論文にもよくみられると指摘されている [16] ように、統計分析の知識や経験がなければ、誤った統計手法の適用や正しく手順をふまないことで間違っただけの解析を行ってしまう恐れがあり、人によっては難しいと思われ統計分析自体を諦めてしまうこともありうる。

また、SPSS などの統計分析専用ソフトは高価でパソコンごとにインストールをする必要がある。オープンソースの R [17] は無料で活用できるがコーディングなどの専門知識を要する。近年、マウスで操作できる「R コマンドー」や「EZR (Easy R)」[18] といったツールも開発されたが、自ら適切かつ順序良く手法を適用、実行しなければなら

ない点は変わらない。

そこで、データ収集 → 比較したい群をドラッグアンドドロップで指定 → 分析実行ボタンをクリック → 示された結果を読み取る、のような簡単な手順で済むシステムの開発が考えられる。これによって、たとえば男女のように比較対照群さえ指定すれば、後述の図 3 や図 5 のように、採用できる検定結果とできない検定結果が視覚的に見分けられる分析結果が得られる。その結果、統計分析の知識や経験がなくても、計算の仕方や数式、公式が分からなくても、コンピュータを使い、コンピュータが示した結果を読み取ることができ、さらに、大人数を対象にした授業でも統計的リサーチリテラシーの素養を結び付ける教育が実施できる。

これらの要求機能を備え付けたシステムはまだなく、実用性の高いシステムの開発が必要であると考えられる。

## 2. 研究目的

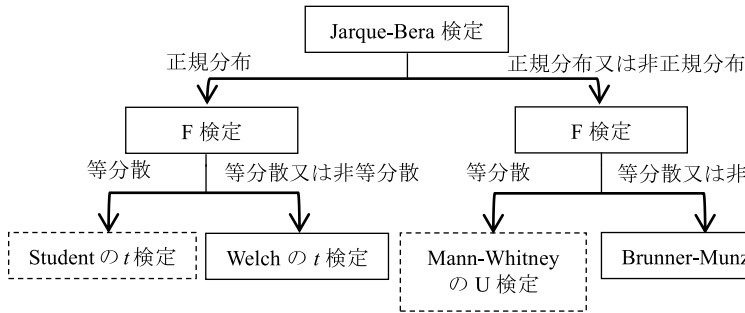
以上より本稿では、統計的検定学習支援用のウェブアプリケーションを開発し、その有用性を検討した。

具体的には、身の回りのデータ収集を容易にする Web 型アンケートフォームの作成と回答機能を備え、得られたデータを用い、統計分析の知識や経験がなくても、計算の仕方や数式、公式が分からなくても、適切かつ手順をふんだ関連のある複数の検定結果が同時に得られ、さらに、採用できる検定結果とできない検定結果が視覚的に見分けられる機能を備える。備え付け検定機能の種類は、正規性検定 (Jarque-Bera 検定)、等分散性の検定 (F 検定、Bartlett 検定)、 $t$  検定 (Student の  $t$  検定と Welch の  $t$  検定、対応のある  $t$  検定)、Mann-Whitney の U 検定、Wilcoxon の符号付順位検定、Brunner-Munzel 検定、一元配置分散分析 (ANOVA)、反復測定による一元配置分散分析 (one way repeated measures ANOVA)、Kruskal-Wallis 検定、Friedman 検定、多重比較 (Tukey-Kramer 法・Steel-Dwass 法・Bonferroni 法・Sidak 法・Holm 法)、カイ二乗検定、Fisher の正確確率検定などである。また、出現確率の  $p$  値は標本数に依存することから実験的操作の効果や変数間の関係の強さを指す指標の効果量 (effect size) も備え付けた ( $r$ ,  $\eta^2$ , partial  $\eta^2$ ,  $f$ ,  $\phi$ , Cramer's  $V$ ) [19]。

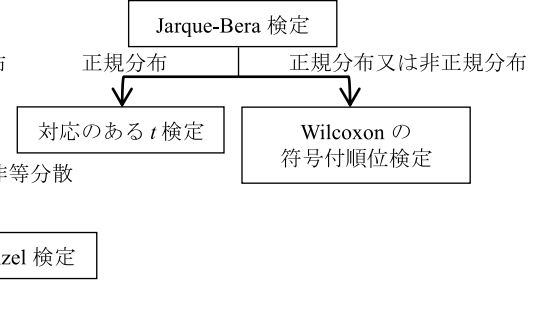
これらの機能を有し、検定結果が Web 上のマウスやボタンクリックだけでリアルタイムで得られるシステムを開発した。その結果、システムの持つ機能によって統計分析の知識や経験がなくても自主的に統計的リサーチリテラシーの素養を身に着けることができる。

統計的検定学習支援用としての本システムの有用性は、大学 3 年生の 100 人を超える 2 回 (2018 年度と 2019 年度) の大人数授業で行い、評価した。具体的には、受講者に本システムを用いたアンケート調査および統計的検定を経験してもらい、その過程で生じた受講者の意識の変容を測ることで確認した。ここでいう受講者の意識とは、統計的検定が

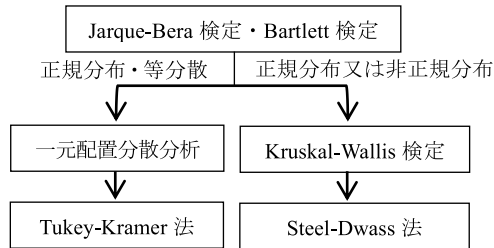
● 対応のない2群



● 対応のある2群



● 対応のない3群以上



● 対応のある3群以上

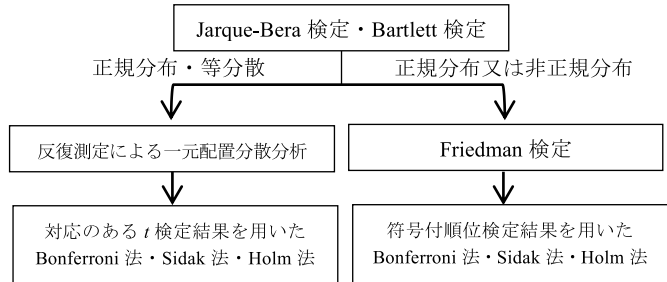


図 1 量的データを用いたシステムの検定フロー  
Fig. 1 Test flow of system using quantitative data.

表 1 開発環境

Table 1 Development environment.

|              |   |
|--------------|---|
| サーバ OS       | Linux   |
| Web サーバ      | Apache  |
| DBMS         | MySQL 5.1.69  |
| 日本語形態素解析エンジン | php-mecab   |
| サーバサイド言語     | PHP Version 5.3.3   |
| クライアントサイド言語  | JavaScript, JQuery, HTML5, CSS3   |
| 動作確認のウェブブラウザ | Chrome 45.0.2454.101 m<br>Internet Explorer 10<br>Firefox 4.0.1.167<br>Safari 6.2<br>iOS/AndroidOS (標準ブラウザ) |

できるかどうかに対する受講者の自己効力感 [20] \*1 を指す。

3. 開発環境

表 1 に開発環境を示した。プログラミング言語は PHP, JavaScript, JQuery, HTML5, CSS3 を、データベースは MySQL を用いた。スマートフォンにも対応している。グラフの作成は Google Charts の API を活用した [21]。統計的検定を行うプログラムなどは自前で組んだ。開発したシステムは <http://www.bumoc.net> サイト上に無料公開されている。

4. 実装した機能の特徴と整合性

本システムの備え付けアンケートフォームの作成と回答

\*1 自己効力感 (Self-efficacy) とは、自分がある状況において必要な行動をうまく遂行できると、自分の可能性を認知していること。また、指示された教育内容を達成していくための行動を、統合し実行する能力を個人的に判断することを学業的な自己効力感という。



図 2 質的データを用いたシステムの検定フロー  
Fig. 2 Test flow of system using qualitative data.

機能を用いて得られるデータの種類の、3段階・4段階・5段階尺度評価による量的データと二者択一・3つ以上の中で1つ選択型・複数選択型・自由記述型による質的データである。そこで、図 1 に量的データを用いた検定フローを、図 2 に質的データを用いた検定フローを示した。

量的データを用いた検定においては、対応のあると対応のないに分け、また、パラメトリック検定とノンパラメトリック検定に分け、それぞれ2群と3群以上の差の検定ができる。3群以上は多重比較の結果も得られる。図 1 をみると、対応の有無および2群と3群以上に分類され、それぞれ非正規分布ならノンパラメトリック検定に分岐、正規分布なら両方とも採用できるためパラメトリック検定にもノンパラメトリック検定にも分岐されている。3群以上の各々の末端枠内の手法は多重比較である。図 2 における→は、期待度数が5未満のマスが出てきたときのみ Fisher の正確確率検定結果が取得できることを指す。

対応の有無の分け方は、たとえば 20 代、30 代、40 代の年齢属性にそれぞれ回答があり、尺度評価の質問項目にもそれぞれの年代の回答者から回答があった場合、「対応のない」2群または3群以上の差の検定画面が表示され実行できる。20代と30代を1つの群にするなどグループ分けが任意で行える。20代のみ回答があった場合は年齢によ

| 項目                  | 計測値A           | 計測値B   |
|---------------------|----------------|--------|
| Data数               | 14             | 11     |
| 合計値                 | 19             | 30     |
| 平均値                 | 1.357          | 2.727  |
| 分散                  | 0.709          | 1.818  |
| JB値                 | 29.629         | 0.495  |
| Jarque-Bera検定       | 非正規分布          | 正規分布   |
| F値                  | 2.565          |        |
| F値出現確率 (p値)         | 0.114          |        |
| 等分散判定               |                | 等分散    |
| Studentのt検定 ×       |                |        |
| 「正規分布に従わない」ため採用できない |                |        |
| Welchのt検定 ×         |                |        |
| 「正規分布に従わない」ため採用できない |                |        |
| Mann-WhitneyのU検定    |                |        |
| 順位和                 | 121.5          | 32.5   |
| 統計量U値               |                | 32.5   |
| 標準偏差                |                | 16.522 |
| 統計量Z値               |                | 2.693  |
| 両側確率 (p値)           |                | 0.007  |
| 差の検定結果              | 1%水準で有意差あり *** |        |
| 効果量(r)と効果量の目安       | 0.539 (効果量大)   |        |
| Brunner-Munzel検定    |                |        |
| 統計量W値               |                | 3.137  |
| 自由度                 |                | 17.683 |
| 両側確率 (p値)           |                | 0.006  |
| 差の検定結果              | 1%水準で有意差あり *** |        |
| 効果量(r)と効果量の目安       | 0.598 (効果量大)   |        |

注) 効果量(r)の目安は、 $r \geq 0.50$  : 大、 $r \geq 0.30$  : 中、 $r \geq 0.10$  : 小、 $r < 0.10$  : ほとんどなし。

図 3 対応のない 2 群間の差の検定結果の表示例

Fig. 3 Example of the independent test results between two groups.

● Mann-Whitney の U 検定

|    |            |       |      |      |         |        |                             |
|----|------------|-------|------|------|---------|--------|-----------------------------|
| 18 | 統計数値表による検定 | 正規化検定 |      |      |         |        |                             |
| 19 | 統計量:U      | 両側検定  | E(U) | V(U) | 連続性補正   | 統計量:z  | 両側P値 * : P<0.05 ** : P<0.01 |
| 20 |            |       |      | 77   | 272.965 | なし     | 2.6934 0.0071**             |
| 21 |            |       |      |      | あり      | 2.6632 | 0.0077**                    |

● Brunner-Munzel 検定

|    |      |        |         |                             |
|----|------|--------|---------|-----------------------------|
| 18 | 検定方法 | 統計量W   | 自由度     | 両側P値 * : P<0.05 ** : P<0.01 |
| 19 | 正規近似 | 3.1375 | -       | 0.0017**                    |
| 20 | 小標本  | 3.1375 | 17.6828 | 0.0058**                    |

図 4 エクセル統計による図 3 の検定結果の整合性確認

Fig. 4 Consistency confirmation of Fig.3 test results by Excel statistics.

る群分けができないため実行不可能の画面に切り替わる。「対応のある」においては、たとえば5段階尺度による質問項目が複数設けられ、2つの質問項目に回答があった場合は2群、3つ以上の質問項目に回答があった場合は対応のある2群と3群以上の差の検定画面が表示される。「対応のある」の群分けは回答者の属性でなく質問項目を用いる。ただ「実験前」、「実験後」のように、各々の質問項目の内容が比較対象になるか否かの判断はユーザに委ねており、そのため、初心者による比較対象を誤った質問項目の選択可能性は否めない。

備え付け検定機能の特徴は、① 関連のある複数の検定が

11.あなたは、bumocを活用すれば、アンケート調査結果で得られたデータを用いて「3群以上の比較の検定」による統計的有意差検定ができると思いますか？  
合計N=82 (n1=39, n2=24, n3=19)

| Jarque-Bera検定  |                |                                  |
|--|----------------|----------------------------------|
| 1群のp値と正規性の判定   | 0.878          | 正規分布                             |
| 2群のp値と正規性の判定   | 0.557          | 正規分布                             |
| 3群のp値と正規性の判定   | 0.518          | 正規分布                             |
| Bartlett検定   |                |                                  |
| M値 (偏り度) とC値 (補正係数)  | 6.489          | 1.019                            |
| $\chi^2$ 値   |                | 6.369                            |
| $\chi^2$ 値の出現確率 (p値)   |                | 0.041                            |
| 等分散判定  |                | 非等分散                             |
| 一元配置分散分析 ×   |                |                                  |
| 「等分散でない」ため採用できない   |                |                                  |
| Kruskal-Wallis検定   |                |                                  |
| 統計量H ( $\chi^2$ 値)   |                | 11.876                           |
| 自由度  |                | 2                                |
| $\chi^2$ 値の出現確率 (p値)   |                | 0.003                            |
| 差の検定結果   | 1%水準で有意差あり *** |                                  |
| 効果量(r)   | 0.381          |                                  |
| Tukey法(Tukey-Kramer)による多重比較 ×  |                |                                  |
| 「等分散でない」ため採用できない   |                |                                  |
| Steel-Dwassa法による多重比較   |                |                                  |
| ステューデント化された箱ひげの表による棄却限界値 (v=∞, 群数=3, aが0.05のq/sqrt(2), aが0.01のq/sqrt(2)) |                | p<0.05 : 2.341<br>p<0.01 : 2.913 |
| 1群と2群の間の統計量  | 0.000          | 有意差無し                            |
| 1群と3群の間の統計量  | 3.471          | 1%水準で有意差あり ***                   |
| 2群と3群の間の統計量  | 2.533          | 5%水準で有意差あり **                    |

注) 効果量 ( $\eta^2$ ) の目安は、 $\eta^2 \geq 0.14$  : 大、 $\eta^2 \geq 0.06$  : 中、 $\eta^2 \geq 0.01$  : 小、 $\eta^2 < 0.01$  : ほとんどなし。効果量 (f) の目安は、 $f \geq 0.40$  : 大、 $f \geq 0.25$  : 中、 $f \geq 0.10$  : 小、 $f < 0.10$  : ほとんどなし。効果量(r)の目安は、 $r \geq 0.50$  : 大、 $r \geq 0.30$  : 中、 $r \geq 0.10$  : 小、 $r < 0.10$  : ほとんどなし。

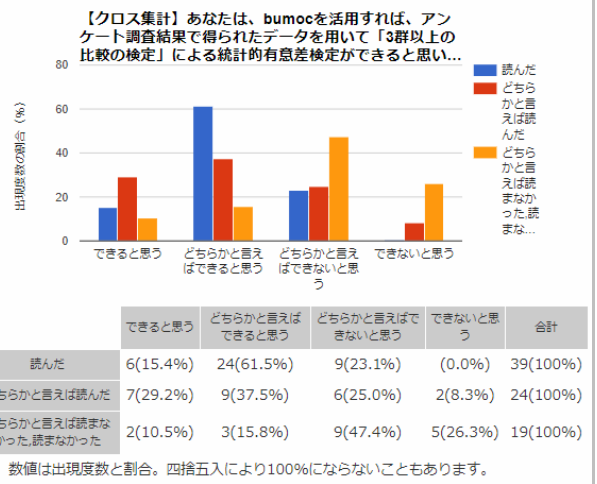


図 5 対応のない 3 群間の差の検定結果の表示例

Fig. 5 Example of the independent test results among three groups.

同時に実行され結果が示されること、また、② 採用してはいけない検定結果にはその理由とともに×マークがつき、ユーザが視覚的に判断できる、ことである。図 3 と図 5 にその例を示した。

図 3 は、対応のない 2 群間の差の検定結果の表示例であり、Jarque-Bera 検定と F 検定、Student の t 検定、Welch の t 検定、Mann-Whitney の U 検定、Brunner-Munzel 検

定および効果量の結果を同時に得た。片方の群の測定値が非正規分布 ( $p < 0.05$ , 以下同様) であることで Student の  $t$  検定と Welch の  $t$  検定は採用できないことを示している。本稿には例を示さなかったが、2つの群の測定値の分散が異なる非等分散 ( $p < 0.05$ , 以下同様) の場合も Student の  $t$  検定は採用できないため×マークが表示される。また、2群間でデータ数が1.5倍以上大きく異なる場合も Student の  $t$  検定は採用できない [22] ように設定している。

なお、本システムでは「対応のない2群」のデフォルトとして、Welch の  $t$  検定と Brunner-Munzel 検定結果のみ示すようにしている。図1の「対応のない2群」における破線の枠はデフォルトの状態では非表示となる検定を意味する。Student の  $t$  検定と Mann-Whitney の U 検定はユーザが同時実行を選択することで図3のように同時に表示される。その理由として、最近では、Student の  $t$  検定は Welch の  $t$  検定にとって代わり使用例が減少の傾向にあり、正規性と等分散性といった前提条件を考慮せずに無条件で Welch の  $t$  検定を採用する方向になっているからである。また、Mann-Whitney の U 検定は非等分散に反応し、状況によっては無視できないほど精度低下を招くことが知られているからである [23]。

図3の例において、正規分布に従わないため、ユーザはノンパラメトリック検定結果のみ採用できることが分かる。一般的な統計学の解説書では、正規性確認について言及せずに  $t$  検定を行うと記載しており、Lang [16] の指摘のように統計学的誤りは研究論文にもよくみられる。正規分布に従わない場合は、 $t$  検定結果はまったく意味のないものなので、本システムにおいては、採用できない検定結果に対してその旨を視覚的に判断でき、ユーザが誤った検定を行うことを防ぐことができる。正規分布の場合は  $t$  検定結果も示され、ノンパラメトリック検定結果と比較できる。

これらの機能によって、前述のように、計算の仕方や数式、公式はあまり重要ではなく、コンピュータを使い、コンピュータが示した結果を読み取ることで、データそのものや統計の活用を意識した統計思考力を養うことが期待でき、その意義は大きい。

図4に、図3の検定結果の整合性を確認するためにエクセル統計を用いた検定結果を示した。図3は図4の検定結果と同様の結果を得ていることが分かる。図3と図4に使用したデータはA群: 1, 2, 1, 1, 1, 1, 1, 1, 1, 2, 4, 1, 1で、B群: 3, 3, 4, 3, 1, 2, 3, 1, 1, 5, 4である。これらのデータはRによる Brunner-Munzel 検定などの結果を例示・説明する際に用いられるインターネット上のものである [24]。統計量  $W$  は 3.137467482303 を四捨五入したもので、図3と図4の統計量  $W$  の違いは表示する桁数が原因である。Lang [16] は、桁数の多い数字は混乱を招くだけで、報告する必要のある最小の  $p$  値は  $p < 0.001$  であるとしており、本システムも表示桁数を小数点3桁にし

ている。本システムに実装したほかの検定機能も R (Ver. 3.5.1) と SPSS (Ver. 21) またはエクセル統計 (BellCurve for Excel Ver. 3.20) などを用いてその整合性を確認した。

図5は後述の表7の2018年度Q4の「実施後」の対応のない3群の差の検定結果であり、Jarque-Bera 検定、Bartlett 検定、一元配置分散分析、Kruskal-Wallis 検定、多重比較 (Tukey-Kramer 法・Steel-Dwass 法) の結果、および効果量を同時に得ている。非等分散であることから一元配置分散分析と Tukey-Kramer 法は採用できないことを視覚的に分かるようになっている。

本稿には示さなかったが、対応のある2群、対応のある3群以上においても関連のある複数の検定結果が同時に示され、採用してはいけない検定に対してはその理由と×マークがつく。質的データを用いた検定では、カイ二乗検定の結果を得られるとともに、期待度数が5未満のマスが出てきたときは Fisher の正確確率検定結果が同時に取得できる。自由記述型による回答は、備え付けのテキストマイニング機能を用いることで形態素解析に加え構文解析やカイ二乗検定と Fisher の正確確率検定を順次行うことができる [25]。効果量もそれぞれ取得できる。加えて、本システムの備え付けアンケート調査機能で得たデータのほか、個別で収集した任意のデータを用いることもできる。図3がその例である。

本システムの具体的なインタフェースや使用方法はマニュアル [26] を参照されたい。

なお、システムの処理速度を示すために、ターンアラウンドタイム [27] \*2を計測した。本システムは、表7の2018年度Q1-Q7の「実施前」と「実施後」の14項目をワンクリックで同時に処理・画面表示する。具体的には、6種類の検定結果を示した図5の処理を14回繰り返して行い、計84の検定結果が同時に得られる。表7は本稿の表の中で最も検定数が多く、処理に負荷がかかる。そこで、表7の2018年度Q1-Q7のすべての検定結果を同時に得るための実行ボタンクリックから画面表示終了までの処理時間を測った。その結果0.18674秒であった。計測時のインターネット回線の下り速度は8.27 MB/secであった。

## 5. システムの有用性評価

### 5.1 方法

大学3年生の情報系演習科目の受講者を対象に、本システムを用いた統計的リサーチリテラシー教育を行い、システムの有用性を検討した。2018年度の受講者数は146人、2019年度は130人である (表2)。システムの持つ学習支援効果を検討するために、授業の実施前と実施後に本

\*2 ターンアラウンドタイム (Turnaround Time) とは、システムに処理要求を送ってから、結果の出力が終了するまでの時間。データやコマンドの入力が終了してから、処理結果の出力が終わって次の要求の受け入れが可能になるまでの時間のこと。

表 3 講義スケジュールと内容  
Table 3 Lecture schedule and contents.

| スケジュール                 | 実施内容   |
|------------------------|--|
| 1 週目<br>所要時間：1 時間      | ①BuMoc を用いた統計的リサーチリテラシー教育授業の「 <b>実施前のアンケート調査</b> 」を実施(回答者数 116)<br>②統計的リサーチリテラシー教育の意義と目的を説明<br>③BuMoc のユーザ登録<br>④BuMoc のアンケート調査フォーム作成機能を活用したアンケート作成の課題提示 (アンケートのテーマは自由設定可, 属性項目数 5 以上, 質問項目数 20 以上必須. 受講者同士でアンケートに回答し合うことで回答数の確保が可能であることも伝えた.)   |
| 2 週目<br>所要時間：1 時間 30 分 | ①プレテスト実施 (各自用意したアンケートを用い, 6 人グループで回答し合う形式で実施した. その際に属性欄には「学籍番号」, 質問項目欄に「本アンケートの修正・改善点を記入してください」を追加させ, 相互回答を促しつつ, 気付いた改善点をコメントし合うようにした. アンケートを用意しなかった人も多数おり, 当日の 24 時まで完成・提出するよう指示した.)<br>②完成したアンケートのタイトルと接続用 URL を当日まで提出を指示した. 提出された接続用 URL は順次ウェブサイト上にリストアップさせ, 相互回答できるようにした.<br>③翌日から 3 週目の授業日まで回答者数 30 以上を確保するように指示(受講登録者数 146 人を 3 つのグループ分けし, 所属グループのメンバーのアンケートに相互回答する形式で回答数を確保).<br>④レポート作成・提出を指示: 提出締切日は 5 週目の授業日まで<br>アンケート調査結果をまとめたレポート作成においては, 検定結果を表と文章でまとめることが必須であることを伝えた. レポートは, 調査背景, 調査目的, 調査方法, 結果と考察, 結論, 感想 (学んだこと, 良かった点と改善点) の順で内容を記述するように指示した.<br>⑤マニュアル活用の促し: 良くわからなくて困ったときはいつでもメールや直接聞きに来て大歓迎だが, 今後の実践に活かすためにもできるだけ各種情報を自らマニュアルを読んで取得するように促した. |
| 3 週目<br>所要時間：30 分      | 回答数が少なかったアンケートに対して 30 分間の相互回答時間を設けた.   |
| 5 週目<br>所要時間：0 分       | レポート提出件数は 111.<br>教員にレポートを提出するとともに, アンケートの分析結果を共有させるために, 互いに回答者宛メールにてレポートのフィードバックを指示した.  |
| 6 週目<br>所要時間：30 分      | ①提出されたレポートの中, 30 分程度で読める分量として 12 本を選定・Web サイト上に公開し, 受講者に閲覧してもらった.<br>②BuMoc を用いた統計的リサーチリテラシー教育授業の「 <b>実施後のアンケート調査</b> 」を実施(回答者数 101)   |

表 2 受講者数と回答者数  
Table 2 Number of students and respondents.

| 対象者           | 2018 年度 |      | 2019 年度 |      |
|---------------|---------|------|---------|------|
|               | n       | %    | n       | %    |
| 受講登録者数        | 146     | 100  | 130     | 100  |
| 実施前のアンケート回答者数 | 116     | 79.5 | 88      | 67.7 |
| 実施後のアンケート回答者数 | 101     | 69.2 | 92      | 70.8 |
| 実施前と後の両方回答者数  | 89      | 61.0 | 61      | 46.9 |

システムを用いてアンケートを行った。実施前と後の両方回答者数はそれぞれの受講登録者の 61.0%と 46.9%であった。本科目の 2018 年度の平均出席率は 72.5%, 2019 年度は 74.8%であり, 表 2 の「実施前のアンケート回答者数」と「実施後のアンケート回答者数」の平均回答率 71.8%は平均出席率の中で得られたものである。

表 3 に 2018 年度の講義スケジュールと実施内容を示した。合計所要時間は約 3 時間 30 分, 所要期間は 6 週間である。実施時間が短く大人数であったこと, また本システムで実施可能な検定は 23 種類 (図 1 と図 2) で多岐にわたること, 座学による検定に関する事前学習は学習意欲を下げってしまうなどの逆効果の恐れがあるなどの理由で, いきなり「自分の問題意識に基づくアンケート作成 → 身近なデータ取得 → 検定実施 → レポートの作成と提出」のスケジュールで実施した。なお, よく分からなくて困ったときはいつ

でもメールや直接聞きに来るように指示するとともに, 各種情報はできるだけ自らマニュアルを読んで取得するように促した。その結果, 2018 年度授業が終了するまでに統計的検定について直接聞きに来た受講者は 0 人であった。

2019 年度にもほぼ同様の形で行った。ただ, 2018 年度には質問項目作成に関する指導を行わなかったが, 2019 年度には 1 週目と 2 週目の間に各自作成・提出したアンケート調査内容を教師が一通りチェックし, 2 週目の授業でフィードバックした。2018 年度の実施後のアンケート調査に「一問一問確認して問題の穴がないかもっと探すべきだった」, 「自分が作ったアンケートより, 他の人のアンケートのできが素晴らしかった」, 「細かい質問項目の変化が重要なのだと思いました」など自由記述の回答があったからである。また, 2019 年度の 6 週目に, 選定したレポート 10 本を閲覧ではなく作成者によるプレゼンテーションをしてもらった (所要時間 90 分)。なお, 2018 年度実施の際に Brunner-Munzel 検定機能が実装されなかったことも異なる点である。2019 年度に統計的検定について直接聞きに来た受講者は 1 人であった。

分析にあたっては, ①自分のデータを用いて統計分析を行う本授業の有意義性を聞き, 全体像を把握した (表 4)。また, 統計的検定の経験による受講者の変容を調べるために「実施前と後の両方回答者」89 人 (2018 年度) と 61 人

表 4 「今回実施したアンケート調査（自分のデータ）による統計分析は、あなたにとって有意義だったか？」に対する評価結果

Table 4 Evaluation results for “Was statistical analysis by the questionnaire conducted this time (using your own data) meaningful for you?”.

| 対象者         | 実施年度 | 区分        | 有意義だった         |        | どちらかといえば有意義だった    |           | 有意義ではなかった |     | 無回答 | 計 |
|-------------|------|-----------|----------------|--------|-------------------|-----------|-----------|-----|-----|---|
|             |      |           | どちらかといえば有意義だった | 有意義だった | どちらかといえば有意義ではなかった | 有意義ではなかった |           |     |     |   |
| 実施後の回答者     | 2018 | n         | 34             | 42     | 15                | 7         | 3         | 101 |     |   |
|             |      | %         | 33.7           | 41.6   | 14.8              | 6.9       | 3.0       | 100 |     |   |
|             |      | 平均得点=3.05 |                |        |                   |           |           |     |     |   |
|             | 2019 | n         | 24             | 57     | 8                 | 2         | 1         | 92  |     |   |
|             |      | %         | 26.1           | 61.9   | 8.7               | 2.2       | 1.1       | 100 |     |   |
|             |      | 平均得点=3.13 |                |        |                   |           |           |     |     |   |
| 実施前と後の両方回答者 | 2018 | n         | 30             | 38     | 13                | 5         | 3         | 89  |     |   |
|             |      | %         | 33.7           | 42.7   | 14.6              | 5.6       | 3.3       | 100 |     |   |
|             |      | 平均得点=3.08 |                |        |                   |           |           |     |     |   |
|             | 2019 | n         | 17             | 37     | 6                 | 1         | 0         | 61  |     |   |
|             |      | %         | 27.9           | 60.7   | 9.8               | 1.6       | 0         | 100 |     |   |
|             |      | 平均得点=3.15 |                |        |                   |           |           |     |     |   |
| 実施後のみ回答者    | 2018 | n         | 4              | 4      | 2                 | 2         | 0         | 12  |     |   |
|             |      | %         | 33.3           | 33.3   | 16.7              | 16.7      | 0         | 100 |     |   |
|             |      | 平均得点=2.83 |                |        |                   |           |           |     |     |   |
|             | 2019 | n         | 7              | 20     | 2                 | 1         | 1         | 31  |     |   |
|             |      | %         | 22.6           | 64.5   | 6.5               | 3.2       | 3.2       | 100 |     |   |
|             |      | 平均得点=3.10 |                |        |                   |           |           |     |     |   |

(2019年度)の回答を分析に用いるが、当該回答者数がそれぞれの受講登録者数の61.0%と46.9%であり、7割強の平均出席率より標本数が少ない。そのため、平均出席率に近い7割前後の「実施後のアンケート回答者数」を「実施前と後の両方回答者」と「実施後のみ回答者」に分けて2群間の差の検定を行い、本稿で用いる分析データの適切性を検討した。② 各々の回答者が本システムを用いた統計的検定の経験で生じた変容(=学習支援効果)を評価するために、実施前と後の両方回答者の回答をそれぞれ用い、4段階評価による対応のある差の検定を実施、システムの影響を検討した(表5)。③ 検定などの統計分析を行った経験の有無による2群分けをし、群間の違いを調べるために群ごとに4段階評価による対応のある差の検定を実施した(表6)。④ 教師による従来型座学の影響を排除しつつ、本システムのみ活用する授業形式をとったことから、統計的検定の手法など受講者にとって主な情報源である備え付けマニュアルの活用度と変容の関連性を分析した(表7)。

なお、後述の検定結果はすべて本システムの備え付け検定機能を活用したものである。また、以下には、関連のある複数の検定結果を同時に示す機能を例示するために複数の検定結果を同時に示した。

## 5.2 結果

表4の「今回実施したアンケート調査（自分のデータ）による統計分析は、あなたにとって有意義だったか？」に対する4段階評価による平均得点は、「実施後の回答者」の2018年度が3.05、2019年度が3.13、「実施前と後の両方回答者」の2018年度が3.08、2019年度が3.15と高い。そして、2018年度の「実施後のみ回答者」を除いて2019年度の「実施後のみ回答者」が3.10で高かった。「有意義だった」、「どちらかといえば有意義だった」の割合の合計は、2018年度は「実施後のみ回答者」を除いてそれぞれ75.3%と76.4%、2019年度は88.0%と88.6%、87.1%であった。

2018年度の「実施後のみ回答者」の平均得点は2.83、「有意義だった」と「どちらかといえば有意義だった」の割合の合計は66.6%で両方とも他より低かった。そこで、2018年度の「実施前と後の両方回答者」と「実施後のみ回答者」に対して差の検定を行った結果、有意差が認められなかった(Brunner-Munzel 検定結果  $p = 0.581$ ,  $r = 0.157$ , Mann-Whitney の U 検定結果  $p = 0.517$ ,  $r = 0.066$  であった。Student の  $t$  検定はデータ数が2群で1.5倍以上異なることや正規分布に従わないため、Welch の  $t$  検定は正規分布に従わないため不採用)。2019年度の「実施前と後の両方回答者」と「実施後のみ回答者」の間にも有意差が認められなかった(Brunner-Munzel 検定結果  $p = 0.773$ ,  $r = 0.037$ , Mann-Whitney の U 検定結果  $p = 0.775$ ,  $r = 0.030$ 。Student の  $t$  検定と Welch の  $t$  検定は上記2018年度と同様)。両年度の「実施前と後の両方回答者」と「実施後のみ回答者」の間に有意差が認められなかったことから、「実施前と後の両方回答者」は本授業の平均出席率とほぼ近い「実施後の回答者」を代表するもの、すなわち本稿で用いる分析データとして適切であると考えられた。

なお、学習にとっては意欲が重要で、学習する達成要求の高い者は、行動の反復によって改良が加えられるが、低い者はそうでないことが実証されている[20]。学習する達成要求の高い者も低い者も混在していると考えられる100人を超えた大人数授業において、上記の平均得点と割合は、本システムを用いた統計的検定の経験が有意義であったことを示唆している。

表5に、授業の「実施前」と「実施後」における受講者の変容を調べる目的で行った4段階評価による対応のある差の検定結果を示した。授業実施後の各々の質問項目にはBuMocを活用すればの語句を追加し、授業実施前の質問項目には当該文章を入れないことで、システムの持つ学習支援



表 5 BuMoc の有用性に関する対応のある差の検定結果  
Table 5 Results of paired test on the effectiveness of BuMoc.

| 質問項目  | 実施年度 | n  | 実施前  | 実施後  | 偏差の<br>平均値 | t検定と<br>Wilcoxonの符<br>号付順位検定 | 効果量                        |
|---|------|----|------|------|------------|------------------------------|----------------------------|
|   |      |    | 平均得点 |      |            |                              |                            |
| Q1: あなたは, BuMoc を活用すれば, 「〇〇の試食による消費者の食嗜好調査」の目的達成のために必要なアンケート調査フォームを作成できると思うか? | 2018 | 86 | 2.58 | 3.05 | 0.47       | $p < 0.001$<br>$p < 0.001$   | $r = 0.408$<br>$r = 0.498$ |
|   | 2019 | 60 | 2.28 | 3.13 | 0.85       | $p < 0.001$<br>$p < 0.001$   | $r = 0.723$<br>$r = 0.843$ |
| Q2: あなたは, BuMoc を活用すれば, Q1 で得られたデータを用い, 基本集計の結果を自分でもまとめることができると思うか?           | 2018 | 85 | 2.41 | 2.84 | 0.43       | $p < 0.001$<br>$p < 0.001$   | $r = 0.407$<br>$r = 0.478$ |
|   | 2019 | 57 | 2.18 | 2.91 | 0.74       | $p < 0.001$<br>$p < 0.001$   | $r = 0.600$<br>$r = 0.680$ |
| Q3: あなたは, BuMoc を活用すれば, Q1 で得られたデータを用いて「2 群の差の検定」による統計的有意差検定ができると思うか?         | 2018 | 86 | 1.96 | 2.91 | 0.95       | $p < 0.001$<br>$p < 0.001$   | $r = 0.664$<br>$r = 0.798$ |
|   | 2019 | 58 | 1.78 | 2.85 | 1.07       | $p < 0.001$<br>$p < 0.001$   | $r = 0.693$<br>$r = 0.759$ |
| Q4: あなたは, BuMoc を活用すれば, Q1 で得られたデータを用いて「3 群以上の差の検定」による統計的有意差検定ができると思うか?       | 2018 | 85 | 1.86 | 2.72 | 0.86       | $p < 0.001$<br>$p < 0.001$   | $r = 0.606$<br>$r = 0.679$ |
|   | 2019 | 59 | 1.56 | 2.63 | 1.07       | ---<br>$p < 0.001$           | ---<br>$r = 0.757$         |
| Q5: あなたは, BuMoc を活用すれば, Q1 で得られたデータを用いて「多重比較」による統計的有意差検定ができると思うか?             | 2018 | 86 | 1.86 | 2.66 | 0.80       | ---<br>$p < 0.001$           | ---<br>$r = 0.686$         |
|   | 2019 | 58 | 1.52 | 2.60 | 1.09       | ---<br>$p < 0.001$           | ---<br>$r = 0.803$         |
| Q6: あなたは, BuMoc を活用すれば, Q1 で得られたデータを用いて「カイ二乗検定」による統計的有意差検定ができると思うか?           | 2018 | 85 | 1.95 | 2.79 | 0.84       | ---<br>$p < 0.001$           | ---<br>$r = 0.618$         |
|   | 2019 | 58 | 1.88 | 2.91 | 1.03       | $p < 0.001$<br>$p < 0.001$   | $r = 0.744$<br>$r = 0.820$ |
| Q7: あなたは, BuMoc を活用すれば, Q1 で得られたデータを用いて「Fisher の正確確率検定」による統計的有意差検定ができると思うか?   | 2018 | 86 | 1.72 | 2.61 | 0.89       | ---<br>$p < 0.001$           | ---<br>$r = 0.677$         |
|   | 2019 | 58 | 1.47 | 2.66 | 1.19       | ---<br>$p < 0.001$           | ---<br>$r = 0.771$         |

効果を評価できるようにした。正規性確認は Jarque-Bera 検定によるもので --- は正規分布に従わないため不採用したことを指す。4段階の尺度は、「できると思う = 4」, 「どちらかといえばできると思う = 3」, 「どちらかといえばできないと思う = 2」, 「できないと思う = 1」, である。

表 5 の Wilcoxon の符号付順位検定結果をみると, Q1 のアンケート調査フォームの作成や Q2 の基本集計, また, Q3 の 2 群の差の検定, Q4 の 3 群以上の差の検定, Q5 の多重比較, Q6 のカイ二乗検定, Q7 の Fisher の正確確率検定によるそれぞれの有意差検定ができるかの質問に対し, すべて高い有意差が認められた ( $p < 0.001$ )。効果量も大きい (効果量の目安は図 5 を参照)。また, それぞれの実施年度において, Q3-Q7 の統計的検定に関わる偏差の平均値は Q1 のアンケート調査フォームの作成と Q2 の基本集計より高い。以上より, 本システムを用いた統計的検定の経験によって受講者の意識に変容が生じたことが確認できた。

なお, Q3-Q7 の質問項目の「~検定ができる」における「できる」の意味には, 言葉の曖昧性は残るが, 「検定結果の読み取りができる」を含んでいるといえる。なぜなら, 本システムは検定結果を自動で示してくれており, 単に  $p$

値のような検定結果を求められるか否かに対する回答であったならば, 表 5 の「実施後」の平均得点は 4.00 または 4 に近い得点になったはずである。ところが, 実際に「できると思う」 (= 4) と回答した人は 2018 年度の Q3 が 22 人, Q4 が 15 人, Q5 が 18 人, Q6 が 18 人, Q7 が 15 人で平均 19.3% しかいなかった。同様に 2019 年度は Q3 が 12 人, Q4 が 10 人, Q5 が 10 人, Q6 が 17 人, Q7 が 13 人, 平均が 20.3% であった。また, 手動で作成する Q1 より自動で検定結果が得られる Q3-Q7 の平均得点が低かった理由も検定結果の読み取りが主な原因であったと考えられるからである。

次に, 回答者の個人差によるシステムの持つ学習支援効果に違いがあるかを確かめるために, 「今回の課題を除いて, アンケート調査で得られたデータを用いてあなた自身が「検定」などの統計分析を行ったことはあるか?」の「ある」群と「ない」群に分け, 授業実施前と後の Q1-Q7 に対し, 群ごとに対応のある差の検定を実施した。表 6 にその結果を示した。Q1-Q7 の詳細と 4 段階の尺度は表 5 と同じであり, --- は非正規分布であったため不採用したことを指す。

Wilcoxon の符号付順位検定結果をみると, 2018 年度の

表 6 検定などの統計分析を行った経験の有無による対応のある差の検定結果

Table 6 Results of paired test depending on existence of experience with statistical analysis such as verification.

| 質問項目 | 実施年度 | 群  | n  | 実施前  | 実施後  | 偏差の<br>平均値 | t 検定と効果量  |           | Wilcoxon の符号付順位検定と<br>効果量 |           |
|------|------|----|----|------|------|------------|-----------|-----------|---------------------------|-----------|
|      |      |    |    | 平均得点 |      |            |           |           |                           |           |
| Q1   | 2018 | ある | 32 | 2.88 | 3.28 | 0.41       | $p=0.062$ | $r=0.328$ | $p=0.092$                 | $r=0.368$ |
|      |      | ない | 54 | 2.42 | 2.89 | 0.46       | $p<0.001$ | $r=0.453$ | $p<0.003$                 | $r=0.573$ |
|      | 2019 | ある | 24 | 2.33 | 3.29 | 0.96       | $p<0.001$ | $r=0.772$ | $p<0.001$                 | $r=0.879$ |
|      |      | ない | 36 | 2.25 | 3.03 | 0.78       | $p<0.001$ | $r=0.688$ | $p<0.001$                 | $r=0.824$ |
| Q2   | 2018 | ある | 31 | 2.74 | 2.97 | 0.23       | $p=0.256$ | $r=0.207$ | $p=0.305$                 | $r=0.230$ |
|      |      | ない | 52 | 2.25 | 2.79 | 0.54       | $p<0.001$ | $r=0.528$ | $p<0.001$                 | $r=0.621$ |
|      | 2019 | ある | 22 | 2.28 | 3.23 | 1.00       | $p<0.001$ | $r=0.707$ | $p=0.001$                 | $r=0.775$ |
|      |      | ない | 35 | 2.14 | 2.71 | 0.57       | $p=0.001$ | $r=0.522$ | $p=0.003$                 | $r=0.598$ |
| Q3   | 2018 | ある | 32 | 2.22 | 3.13 | 0.91       | $p<0.001$ | $r=0.667$ | $p<0.001$                 | $r=0.822$ |
|      |      | ない | 52 | 1.85 | 2.79 | 0.94       | $p<0.001$ | $r=0.651$ | $p<0.001$                 | $r=0.784$ |
|      | 2019 | ある | 23 | 1.96 | 2.91 | 0.96       | $p=0.001$ | $r=0.613$ | $p=0.005$                 | $r=0.646$ |
|      |      | ない | 35 | 1.66 | 2.80 | 1.14       | $p<0.001$ | $r=0.747$ | $p<0.001$                 | $r=0.842$ |
| Q4   | 2018 | ある | 33 | 2.00 | 2.97 | 0.97       | $p<0.001$ | $r=0.648$ | $p<0.001$                 | $r=0.729$ |
|      |      | ない | 50 | 1.80 | 2.56 | 0.76       | $p<0.001$ | $r=0.560$ | $p<0.001$                 | $r=0.634$ |
|      | 2019 | ある | 23 | 1.61 | 2.78 | 1.17       | ---       | ---       | $p=0.002$                 | $r=0.690$ |
|      |      | ない | 36 | 1.53 | 2.53 | 1.00       | $p<0.001$ | $r=0.717$ | $p<0.001$                 | $r=0.831$ |
| Q5   | 2018 | ある | 32 | 1.97 | 2.84 | 0.88       | $p<0.001$ | $r=0.629$ | $p<0.001$                 | $r=0.762$ |
|      |      | ない | 52 | 1.83 | 2.60 | 0.77       | ---       | ---       | $p<0.001$                 | $r=0.641$ |
|      | 2019 | ある | 23 | 1.57 | 2.61 | 1.04       | $p<0.001$ | $r=0.791$ | $p<0.001$                 | $r=0.878$ |
|      |      | ない | 35 | 1.49 | 2.60 | 1.11       | $p<0.001$ | $r=0.724$ | $p<0.001$                 | $r=0.771$ |
| Q6   | 2018 | ある | 31 | 2.19 | 2.90 | 0.71       | $p=0.009$ | $r=0.453$ | $p=0.016$                 | $r=0.493$ |
|      |      | ない | 52 | 1.85 | 2.73 | 0.89       | ---       | ---       | $p<0.001$                 | $r=0.671$ |
|      | 2019 | ある | 23 | 2.13 | 3.22 | 1.09       | ---       | ---       | $p<0.001$                 | $r=0.879$ |
|      |      | ない | 35 | 1.71 | 2.71 | 1.00       | $p<0.001$ | $r=0.734$ | $p<0.001$                 | $r=0.788$ |
| Q7   | 2018 | ある | 32 | 1.94 | 2.88 | 0.94       | $p<0.001$ | $r=0.608$ | $p=0.001$                 | $r=0.688$ |
|      |      | ない | 52 | 1.62 | 2.48 | 0.87       | ---       | ---       | $p<0.001$                 | $r=0.667$ |
|      | 2019 | ある | 23 | 1.48 | 2.91 | 1.44       | ---       | ---       | $p=0.001$                 | $r=0.776$ |
|      |      | ない | 35 | 1.46 | 2.49 | 1.03       | $p<0.001$ | $r=0.698$ | $p<0.001$                 | $r=0.777$ |

「ある」の Q1 のアンケート調査フォームの作成と Q2 の基本集計を除くすべての項目で有意差が認められ、効果量も 2018 年度の Q6 が中で他は大きかった。検定機能と無関係の質問項目の 2018 年度の Q1 と Q2 を除くすべての統計的検定機能において、個人差に関係なく本システムの活用による受講者の意識に変容が生じたことが確認できた。

表 7 に「BuMoc のマニュアルを読んだか？」に対して「読んだ (1 群)」と「どちらかといえば読んだ (2 群)」, 「どちらかといえば読まなかった・読まなかった (3 群)」の 3 群分けによる差の検定結果を示した。質問項目の詳細は表 5 と同じであり、Jarque-Bera 検定結果 2019 年度の Q7 を除くすべてが正規分布であった。等分散性の確認は Bartlett 検定 ( $p < 0.05$ ) で行った。差の検定は一元配置分散分析と Kruskal-Wallis 検定で、多重比較は Tukey 法 (Tukey-Kramer) と Steel-Dwass 法で行った。多重比較の有意差判定はステューデント化された範囲の表による棄却限界値を適用した。-- は、非等分散であったため不採用したことを指す。- は、Kruskal-Wallis 検定で有意でなかったため多重比較を行わなかったことを指す。平均得点は 4 段階評価によるものであり、尺度は、「読んだ = 4」, 「どちらかといえば読んだ = 3」, 「どちらかといえば読まなかった

= 2」, 「読まなかった = 1」, である。

表の 2019 年度の Q3 の平均得点において、1 群は「実施前」1.18 から「実施後」3.19, 2 群は 1.72 から 2.59, 3 群は 1.75 から 2.53 へ変容した。すなわち実施前には平均得点が最も低かった 1 群が、マニュアルを読むことで実施後に他群より高く変容したことになる。2018 年度の Q4 と 2019 年度の Q5 も同じであった。また、「実施前」の 1 群の平均得点が他群より高かった 2018 年度の Q1, 2018 年度の Q2, 2018 年度の Q3, 2019 年度の Q6 を除くほかも「実施後」に他群より平均得点が最も高く変容した。加えて、2018 年度の Q7 を除く「実施後」のすべての 1 群の平均得点は他より高い。

Kruskal-Wallis 検定と Steel-Dwass 法による多重比較の結果をみると、「実施前」は 1 群と他群間に違いは認められなかったが、「実施後」の Q3 と Q5 に両年度ともに有意差が認められた。また、Q6 を除く Q1 と Q2, Q4 は 2018 年度に、Q7 は 2019 年度に 1 群と他群間に有意差が認められた。

以上から、本システムを用いた統計的検定の学習にあたって主な情報源であるマニュアルの活用は有効であることが示唆された。

表 7 「BuMoc のマニュアルを読んだか？」に対する「読んだ (1 群)」と「どちらかといえば読んだ (2 群)」, 「どちらかといえば読まなかった・読まなかった (3 群)」における授業実施前と後の 3 群間の差の検定結果

Table 7 Test results of the difference among 3 groups before and after the class in response to the question “Did you read the BuMoc’s manual?” (Group 1: read, group 2: somewhat read and group 3: rather did not read and did not read).

| 質問項目 | 実施年度 | 群                 | 実施前            |                      |                       |   |  | 実施後            |                      |                       |   |   |
|------|------|-------------------|----------------|----------------------|-----------------------|---|--|----------------|----------------------|-----------------------|---|---|
|      |      |                   | n              | 平均得点                 | Bartlett 検定による等分散性の検定 | 一元配置分散分析と効果量 Kruskal-Wallis と効果量                      | ①Tukey 法 (Tukey-Kramer) ②Steel-Dwass 法 | n              | 平均得点                 | Bartlett 検定による等分散性の検定 | 一元配置分散分析と効果量 Kruskal-Wallis と効果量                      | ①Tukey 法 (Tukey-Kramer) ②Steel-Dwass 法  |
| Q1   | 2018 | 1 群<br>2 群<br>3 群 | 39<br>24<br>21 | 2.67<br>2.46<br>2.57 | $p=0.809$             | $p=0.664$<br>$\eta^2=0.010$<br>$p=0.602$<br>$r=0.110$ | ①有意差無<br>② -                           | 40<br>23<br>20 | 3.20<br>3.13<br>2.50 | $p=0.922$             | $p=0.004$<br>$\eta^2=0.127$<br>$p=0.005$<br>$r=0.360$ | 多重比較①と②による 1 群と 3 群間 ( $p<0.01$ ), 2 群と 3 群間 ( $p<0.05$ )                                 |
|      | 2019 | 1 群<br>2 群<br>3 群 | 27<br>18<br>16 | 2.33<br>2.39<br>2.13 | $p=0.821$             | $p=0.590$<br>$\eta^2=0.018$<br>$p=0.589$<br>$r=0.132$ | ①有意差無<br>② -                           | 27<br>18<br>15 | 3.26<br>2.94<br>3.13 | $p=0.378$             | $p=0.315$<br>$\eta^2=0.040$<br>$p=0.363$<br>$r=0.184$ | 多重比較①と②による有意差無  |
| Q2   | 2018 | 1 群<br>2 群<br>3 群 | 39<br>24<br>21 | 2.54<br>2.46<br>2.14 | $p=0.148$             | $p=0.153$<br>$\eta^2=0.045$<br>$p=0.118$<br>$r=0.226$ | ①有意差無<br>② -                           | 39<br>23<br>20 | 3.05<br>2.91<br>2.35 | $p=0.048$             | --<br>--<br>$p=0.008$<br>$r=0.345$                    | 多重比較②による 1 群と 3 群間 ( $p<0.01$ )   |
|      | 2019 | 1 群<br>2 群<br>3 群 | 27<br>18<br>15 | 2.19<br>2.44<br>1.93 | $p=0.169$             | $p=0.221$<br>$\eta^2=0.052$<br>$p=0.273$<br>$r=0.208$ | ①有意差無<br>② -                           | 27<br>16<br>15 | 3.15<br>2.75<br>2.73 | $p=0.426$             | $p=0.117$<br>$\eta^2=0.075$<br>$p=0.139$<br>$r=0.261$ | 多重比較①と②による有意差無  |
| Q3   | 2018 | 1 群<br>2 群<br>3 群 | 39<br>24<br>21 | 2.05<br>1.96<br>2.00 | $p=0.023$             | --<br>--<br>$p=0.980$<br>$r=0.022$                    | ① --<br>② -                            | 39<br>24<br>20 | 3.05<br>3.04<br>2.45 | $p=0.257$             | $p=0.016$<br>$\eta^2=0.098$<br>$p=0.021$<br>$r=0.306$ | 多重比較①による 1 群と 3 群間 ( $p<0.05$ ) と 2 群と 3 群間 ( $p<0.05$ ), 多重比較②による 1 群と 3 群間 ( $p<0.05$ ) |
|      | 2019 | 1 群<br>2 群<br>3 群 | 27<br>18<br>16 | 1.18<br>1.72<br>1.75 | $p=0.989$             | $p=0.930$<br>$\eta^2=0.002$<br>$p=0.916$<br>$r=0.054$ | ①有意差無<br>② -                           | 26<br>17<br>15 | 3.19<br>2.59<br>2.53 | $p=0.578$             | $p=0.014$<br>$\eta^2=0.144$<br>$p=0.020$<br>$r=0.367$ | 多重比較①による 1 群と 2 群間 ( $p<0.05$ ) と 1 群と 3 群間 ( $p<0.05$ ), 多重比較②による 1 群と 2 群間 ( $p<0.05$ ) |
| Q4   | 2018 | 1 群<br>2 群<br>3 群 | 39<br>24<br>21 | 1.82<br>1.92<br>1.90 | $p=0.247$             | $p=0.877$<br>$\eta^2=0.003$<br>$p=0.827$<br>$r=0.067$ | ①有意差無<br>② -                           | 39<br>24<br>19 | 2.92<br>2.88<br>2.11 | $p=0.041$             | --<br>--<br>$p=0.003$<br>$r=0.381$                    | 多重比較②による 1 群と 3 群間 ( $p<0.01$ ) と 2 群と 3 群間 ( $p<0.05$ )                                  |
|      | 2019 | 1 群<br>2 群<br>3 群 | 27<br>18<br>16 | 1.56<br>1.44<br>1.81 | $p=0.176$             | $p=0.380$<br>$\eta^2=0.033$<br>$p=0.568$<br>$r=0.136$ | ①有意差無<br>② -                           | 27<br>17<br>15 | 2.89<br>2.41<br>2.40 | $p=0.955$             | $p=0.114$<br>$\eta^2=0.075$<br>$p=0.131$<br>$r=0.263$ | 多重比較①と②による有意差無  |
| Q5   | 2018 | 1 群<br>2 群<br>3 群 | 39<br>24<br>21 | 1.85<br>2.04<br>1.81 | $p=0.144$             | $p=0.642$<br>$\eta^2=0.011$<br>$p=0.740$<br>$r=0.085$ | ①有意差無<br>② -                           | 39<br>24<br>20 | 2.92<br>2.79<br>2.10 | $p=0.321$             | $p=0.002$<br>$\eta^2=0.142$<br>$p=0.004$<br>$r=0.368$ | 多重比較①による 1 群と 3 群間 ( $p<0.01$ ) と 2 群と 3 群間 ( $p<0.05$ ), 多重比較②による 1 群と 3 群間 ( $p<0.01$ ) |
|      | 2019 | 1 群<br>2 群<br>3 群 | 27<br>17<br>16 | 1.56<br>1.24<br>1.75 | $p=0.058$             | $p=0.099$<br>$\eta^2=0.078$<br>$p=0.116$<br>$r=0.268$ | ①有意差無<br>② -                           | 27<br>17<br>15 | 2.89<br>2.12<br>2.53 | $p=0.723$             | $p=0.021$<br>$\eta^2=0.129$<br>$p=0.028$<br>$r=0.348$ | 多重比較①による 1 群と 2 群間 ( $p<0.05$ ), 多重比較②による 1 群と 2 群間 ( $p<0.05$ )                          |
| Q6   | 2018 | 1 群<br>2 群<br>3 群 | 39<br>24<br>21 | 1.97<br>2.04<br>1.81 | $p=0.433$             | $p=0.749$<br>$\eta^2=0.007$<br>$p=0.822$<br>$r=0.068$ | ①有意差無<br>② -                           | 38<br>24<br>20 | 2.95<br>2.88<br>2.45 | $p=0.537$             | $p=0.095$<br>$\eta^2=0.058$<br>$p=0.089$<br>$r=0.243$ | 多重比較①と②による有意差無  |
|      | 2019 | 1 群<br>2 群<br>3 群 | 27<br>18<br>15 | 2.07<br>1.67<br>1.67 | $p=0.713$             | $p=0.191$<br>$\eta^2=0.056$<br>$p=0.218$<br>$r=0.225$ | ①有意差無<br>② -                           | 27<br>17<br>15 | 3.15<br>2.65<br>2.73 | $p=0.706$             | $p=0.156$<br>$\eta^2=0.064$<br>$p=0.151$<br>$r=0.253$ | 多重比較①と②による有意差無  |
| Q7   | 2018 | 1 群<br>2 群<br>3 群 | 39<br>24<br>21 | 1.74<br>1.96<br>1.57 | $p=0.004$             | --<br>--<br>$p=0.573$<br>$r=0.115$                    | ① --<br>② -                            | 39<br>24<br>20 | 2.72<br>2.88<br>2.25 | $p=0.957$             | $p=0.053$<br>$\eta^2=0.071$<br>$p=0.039$<br>$r=0.279$ | 多重比較①による有意差無, 多重比較②による 2 群と 3 群間 ( $p<0.05$ )   |
|      | 2019 | 1 群<br>2 群<br>3 群 | 27<br>18<br>16 | 1.44<br>1.33<br>1.56 | 正規分布に従わないため等分散確認は不要   | 正規分布に従わないため採用できない<br>$p=0.779$<br>$r=0.090$           | ①正規分布に従わないため採用できない<br>② -              | 27<br>16<br>15 | 3.11<br>2.06<br>2.47 | $p=0.766$             | $p=0.001$<br>$\eta^2=0.222$<br>$p=0.002$<br>$r=0.456$ | 多重比較①による 1 群と 2 群間 ( $p<0.01$ ), 多重比較②による 1 群と 2 群間 ( $p<0.01$ )                          |

5.3 今後の課題

まずは、「マニュアルの説明文に内容の意味が難しい箇所があったのもう少しかみ砕いて書いてほしい。」とい

う自由記述の回答があり, マニュアルの有用性をさらに引き上げるためにはマニュアルのよりきめ細かな記述が望まれた。

また、「アンケート項目の作成，集計などが容易にできる．アンケート回答先，集計，統計が1つのサイトで一貫して行える」，「分析ツールがすでにできあがっているので，操作しやすく，面白いと思う」といった肯定的な自由記述の回答もあるものの，「操作しにくい」，「使い方が難しい」，「やり方が分かりにくい」といった否定的な回答もみられたことからシステムの使いやすさもさらに改善する必要がある．

なお，大学の15コマ授業で表3の授業形式を2度実施することで，統計的検定の経験を繰り返した際の学習支援効果を評価することも有効であろうと考える．受講者は自らの失敗から学び，また1度目の経験から生じた変容を生かすことができ，より高い変容が期待されるからである．

## 6. 議論

### 6.1 学術的・技術的意義

本稿の学術的・技術的意義は，① 統計的検定手法の選択の誤りへの対応，② 統計的検定の簡便さをあげられる．

① 統計的検定手法の選択の誤りへの対応：前述のようにデータが正規分布に従わないときに，パラメトリック検定を使用してしまう統計学的誤りは研究論文にもよくみられる [16]．本システムは採用できない検定手法に対して自動判別機能を備え付けており，統計的検定手法の選択の誤りを防止できる．

② 統計的検定の簡便さ：本システムとエクセル統計を用いて表7の2018年度の実施前と実施後の一元配置分散分析とKruskal-Wallis検定結果を得るまでのプロセスを比較すると，本システムは1回の実行クリックで結果を得られる．エクセル統計は28回繰り返して実行しなければならない．また，エクセル統計は複数の分析ツールの中で自ら適切かつ順序良く手法を適用，実行しなければならない．本システムの簡便さによって，平均値や比率の比較，クロス集計などとどまりがちなデータ分析を，統計的検定まで繰り返し経験でき，統計の活用を意識した統計思考力を養うことが期待できる．また，手間も省けて複数の統計的検定結果を同時に検討できることで有意な差を見抜くことができる効果も期待される．エクセル統計を比較対象とした理由は，SPSSなどに比べ精度に遜色なく手順が簡便で費用の面からも教育現場に導入しやすいからである．

### 6.2 初心者の学習支援用としての限界

前述した本システムの持つ学習支援効果は回答者の「意識」を問い評価したもので，実際に正しい統計的検定ができるか否かに関する「事実」を評価したものではない．本システムは正しい統計的検定結果を自動で示してくれることを前提に議論を進めているため，ここでいう事実とは，示された検定結果を正しく解釈できるか否かを指す．

そこで，星野 [28] によると，諸要因が制御された実験

データでない調査観察データによる因果推論を行う際には，データ偏りや選択バイアスを無視してしまうと誤った結論が導かれることすらある．すなわち，備え付けアンケート調査機能を用いることで得られる観察データの分析において，特に初心者による誤った解釈が行われてしまう可能性は否めない．既存の統計ソフトも同様であるが，本システムも，誤った解釈を防げるような機能はまだない．現状では文献や教師がその役割を果たしており，正しい解釈支援機能の開発とその有用性評価は残された今後の課題である．

他方，学部生による本システムを用いたアンケート調査および統計的検定結果が学術論文として採択された事例がある [29]．事例の論文仕上げにおいて本システムのマニュアルが参照されつつ教師の手助けが加わった．初心者による本システムの活用の際には誤った解釈を防げるためのサポートが望まれる．

### 6.3 既存統計ソフトとの比較調査の限界

SPSSやエクセル統計，Rなどの既存の統計ソフトを利用した場合と本システムを利用した場合を比較した調査も検討したが，本稿では比較調査を行っていない．その理由は，実験的に比較するまでもなく，次の点で本システムに実装した検定機能が優れているからである．既存の統計ソフトを用いる場合は，複数の分析ツールの中で自ら適切かつ順序良く手法を適用，実行しなければならない．しかし，本システムはこれらの自動対応機能が備えられているため，比較調査を行うと，自動でやってくれる本システムの評価が高くなるのは想像に難くない．この点について「(BuMocは)統計が1つのサイトで一貫して行える．分析ツールがすでにできあがっているので，操作しやすく，面白いと思う」といった自由記述の回答からも裏付けられる．

### 6.4 受講者の変容要因の区分の限界

本研究において，表3のように教師による従来型座学の影響を排除した講義スケジュールの適用や，本システムとマニュアルのみ活用による自主的な学習形式をとったこと，また，「2018年度」と「2019年度」の授業期間中に，統計的検定について教師に聞きに来た受講者は1人しかいなかったことなどから，授業による受講者への影響は極力排除できるよう実験をデザインした．これらをふまえたうえで，本システムの起因による変容のみを測る目的で授業実施後のアンケート質問項目には BuMocを活用すればの語句を追加し，データを取得，比較分析した．ところが，表5の「偏差の平均値」をみると，2018年度より2019年度のほうが全項目で高く，その理由として，2018年度とは違って，2019年度にはアンケート調査内容を教師が一通りチェックしたこと，選定されたレポートの閲覧でなく作成者にプレゼンテーションしてもらった形で分析事例を共有したことが変容要因として影響した可能性もうかがえる．し

かし、本稿の調査では授業による受講者の変容要因を明確に区分、提示できずこの点について限りがある。

## 6.5 分析に用いたデータの限界

本稿では「実施前と後の両方回答者」のデータを分析に用いており、受講者の一部からしか回答は得られていない。そのため、本システムを用いた授業に対して、学習する達成要求の低い者の回答をより多く含んだときに本稿と同様の結果が得られるとは限らない。したがって、本稿の結果の一般化には留意する必要がある。

## 7. おわりに

自由記述回答の「(BuMoc は) アンケート項目の作成、集計などが容易にできる。アンケート回答先、集計、統計が1つのサイトで一貫して行える」が示唆するように、統計分析の知識や経験がなくても、計算の仕方や数式、公式が分からなくても、適切かつ手順をふんだ関連のある複数の検定結果が同時に得られ、また、採用できる検定結果とできない検定結果が視覚的に見分けられる機能を備えた統計的リサーチリテラシー教育支援用システムを開発した。本システムを用いることで、比較したい群をドラッグアンドドロップで指定するだけで正規性の検定から多重比較まで23種類の検定結果が得られる。

受講者数100人を超える大学3年生の授業でシステムの持つ学習支援効果を検討した。授業実施前と後における対応のある差の検定結果、有意差が認められた。ユーザはシステムのマニュアルを読むことで自主的に各種検定手法を学ぶことができること、また、検定などの統計分析を行った経験の有無にかかわらず学習支援効果が高いことが示唆された。

今後の課題として、システムのインタフェースの向上とマニュアルの充実および正しい解釈支援機能の開発が残された。

謝辞 この成果は科学研究費補助金(基盤研究(C):課題番号25450329)の助成に基づいたものである。

## 参考文献

- [1] 文部科学省中央教育審議会: 21世紀の知識創造社会に向けた統計教育推進への要望書, 日本統計学会誌, Vol.36, No.2, pp.349-358 (2007).
- [2] 垣花京子: 不確実性事象に対する統計リテラシー教育に関する一考察, 筑波学院大学紀要, No.10, pp.1-10 (2015).
- [3] 林 創, 山田剛史: リサーチリテラシーの育成による批判的思考態度の向上—「書く力」と「データ分析力」を中心に, 京都大学高等教育研究, No.18, pp.41-51 (2012).
- [4] 谷岡一郎: データはウソをつく—科学的な社会調査の方法, ちくまプリマー新書, p.169 (2007).
- [5] 片瀬一男: 情報化社会における市民的教養教育としての社会調査教育—統計的リサーチ・リテラシーの育成を中心に, 社会学評論, Vol.58, No.4, pp.476-491 (2008).
- [6] Cobb, G.W.: Teaching statistics, in *Heeding the Call for*

- Change: Suggestions for Curricular Action (MAA Notes No.22), Edited by Lynn A. Steen, pp.3-43 (1992).
- [7] 渡辺美智子: 知識基盤社会における統計教育の新しい枠組み: 科学的探究・問題解決・意思決定に至る統計思考力, 日本統計学会誌, Vol.42, No.2, pp.253-271 (2013).
- [8] 渡辺美智子: 教養教育としての統計とデータサイエンス教育の課題—意思決定を高度化する統計思考力の育成, 東北大学高度教養教育・学生支援機構紀要, Vol.2, pp.31-39 (2016).
- [9] 瀬沼花子: 企業の算数・数学教育への期待: データに基づく予測の強調と指導法の改善, 科学教育研究, pp.34-42 (2004).
- [10] 学習指導要領「生きる力」: 文部科学省, 入手先 ([http://www.mext.go.jp/a\\_menu/shotou/new-cs/youryou/1304449.htm](http://www.mext.go.jp/a_menu/shotou/new-cs/youryou/1304449.htm)) (参照 2019-08-01).
- [11] 統計教育の場を通じた統計リテラシー等の向上に向けて: 総務省, 入手先 ([http://www.soumu.go.jp/main\\_content/000514026.pdf](http://www.soumu.go.jp/main_content/000514026.pdf)) (参照 2019-08-01).
- [12] 渡辺美智子, 末永勝征, 竹内光悦, 宿久 洋, 山口和範, 浅野長一郎: インターネット環境における統計科学普及のための教育用サイト構築の試み, 統計数理, Vol.49, No.2, pp.241-260 (2001).
- [13] 渡辺美智子, 末永勝征, 熊原啓作: マルチメディアを活用した統計教育の情報化に関する研究, 放送大学研究年報, No.25, pp.117-126 (2007).
- [14] おしゃべりな部屋 (プラネタリウム, 星, 植物, 熱帯魚, 統計学), 入手先 (<http://aoki2.si.gunma-u.ac.jp>) (参照 2019-08-01).
- [15] Google フォーム-アンケートを作成、分析できる無料サービス, 入手先 (<https://www.google.com/intl/ja-jp/forms/about/>) (参照 2019-08-01).
- [16] Lang, T.: Twenty Statistical Errors Even YOU Can Find in Biomedical Research Articles, *Croatian Medical Journal*, Vol.45, No.4, pp.361-370 (2004).
- [17] R: The R Project for Statistical Computing, 入手先 (<https://www.r-project.org>) (参照 2019-08-01).
- [18] Kanda, Y.: Investigation of the freely available easy-to-use software 'EZ'R' for medical statistics, *Bone Marrow Transplantation*, Vol.48, pp.452-458 (2013).
- [19] 水本 篤, 竹内 理: 研究論文における効果量の報告のために—基礎的概念と注意点, 英語教育研究, Vol.31, pp.56-66 (2008).
- [20] アルバート・バンデューラ (編), 本明 寛ほか (訳): 激動社会の中の自己効力, 金子書房, p.352 (1997).
- [21] Google Charts, 入手先 (<https://google-developers.appspot.com/chart/>) (参照 2019-02-20).
- [22] 秋山 徹: バイオ実験に絶対使える統計の基本 Q&A, 羊土社, p.254 (2012).
- [23] 名取真人: マン・ホイットニーの U 検定と不等分散時における代表値の検定法, 霊長類研究, Vol.30, No.1, pp.173-185 (2014).
- [24] Brunner-Munzel 検定, 入手先 (<https://oku.edu.mie-u.ac.jp/~okumura/stat/brunner-munzel.html>) (参照 2019-08-01).
- [25] 朴 壽永, 長谷部正, 安江絳幸: ウェブ型テキストマイニングツール iTM の開発, システム農学, Vol.32, No.1, pp.25-35 (2012).
- [26] BuMoc マニュアル, 入手先 ([http://www.bumoc.net/document/bumoc\\_manual.pdf](http://www.bumoc.net/document/bumoc_manual.pdf)) (参照 2019-08-01).
- [27] ターンアラウンドタイム (Turn Around Time) とは: IT用語辞典, 入手先 (<http://e-words.jp/w/ターンアラウンドタイム.html>) (参照 2019-08-01).
- [28] 星野崇宏: 調査観察データの統計科学—因果推論・選択バイアス・データ融合, 岩波書店, p.245 (2009).
- [29] 小田恭平, 新部昭夫, 朴 壽永: 農業分野におけるクラ

ウドファンディングの活用現状と成功要因，農業情報研究，Vol.28, No.2, pp.86-96 (2019).



朴 壽永 (正会員)

2009年東京農業大学大学院農学研究科博士課程修了。同年東京情報大学博士研究員。東北大学大学院農学研究科・農学部特任助教を経て，2016年東京農業大学国際食料情報学部助教。

2018年准教授。2020年より県立広島大学生物資源科学部教授。博士（国際バイオビジネス学）。2014年システム農学会奨励賞受賞。



山田 崇裕

2010年東京農業大学大学院農学研究科博士後期課程修了。東京農業大学特別研究員を経て，2011年東京農業大学国際食料情報学部助教。2017年より准教授。博士（国際バイオビジネス学）。



新部 昭夫

1983年東京農業大学大学院農学研究科農学専攻博士後期課程修了。同年東京農業大学農学部電算機室副手を経て，1986年同助手，1990年同講師，1998年東京農業大学国際食料情報学部助教授，2008年より教授。農学博士。

専門は動物遺伝学，農業情報処理。作物モデリングも行う。



安江 紘幸

2008年東京農業大学大学院農学研究科博士課程修了。同年東京農業大学国際バイオビジネス学科助手。2009年東北大学大学院農学研究科・農学部助教を経て2013年国立研究開発法人農業・食品産業技術総合研究機構東北農業研究センター研究員。2018年より主任研究員。博士（国際バイオビジネス学）。2016年東北農業経済学会木下賞（奨励賞）受賞。



井形 雅代

1989年東京農業大学卒業。1990年同農業経済経済学科助手。1998年同生物企業情報学科（国際バイオビジネス学科）講師，2004年准教授。2004～2005年オランダワーヘニンゲン大学客員研究員。