

# 独立性の検定に関するHTMLをベースとした教材の 操作ログの解析

山口大学・教育学部 北本 卓也

Takuya Kitamoto, Faculty of Education, Yamaguchi University

東邦大学・薬学部 金子 真隆

Masataka Kaneko, Faculty of Pharmaceutical Sciences, Toho University

東邦大学・理学部 野田 健夫

Takeo Noda, Faculty of Science, Toho University

## 概要

We will propose an HTML-based teaching material on "Tests of Independence," which has a wide range of applications in the field of medical statistics, etc. Then, we will explain how to utilize the teaching material for use in large classes. Furthermore, we will discuss how to provide efficient learning support by matching the learner's operation logs during the class with the results of the post-test.

## 1 はじめに

筆者の先行研究 [1] において、医療統計分野などで応用範囲の広い「独立性の検定」に関するHTMLをベースとした教材について提案し、これを少人数形式の授業で利用した際の学習状況の分析に基づいて改変していく経緯について説明した。経験上、この学習トピックは学習者にかかる認知負荷 [2] が高く、これをいかに低減するかという点がポイントであった。本稿では、以上の流れから出来上がった統合的な教材と、これを大人数の授業で利用するための授業設計について説明した上で、授業時の学習者の操作ログと事後テストの結果を照合することにより、効率的な学習支援を行う方策について検討する。少人数を対象とした授業であれば、先行研究 [1] の事例にもあるように、think aloudに近い作業フローを採用することで、学習者の思考状況を追跡することが可能だが、通常の大人数を対象とした授業、特にオンラインでの授業の場合にはこれは困難である。本研究のアプローチは、操作中の状況をモニターする段階から、at riskの学習者を特定すると同時に、有効な教育的介入の方策を見出すことを目指している。もとより、操作ログ単体で学習者にかかる認知負荷を十分な程度に把握するのは容易でないことが想定されるが、アプリケーションの導入に様々な制約が加わる学校現場において、教授者が学習者の状況を的確に把握しながら探究的な数理科学の学びを進める上で、このような方策は不可避なものになると考えられる。

## 2 HTML をベースとした教材について

本教材は HTML 上の教材であり、ブラウザ上で動作する。このため、パソコンのみならず、タブレットやスマートフォンから閲覧することが可能であり、教材としての汎用性が高い。

また、本教材は [3] のシステムを用いて構成されている。この [3] のシステムでは、その構成をブラウザ上から行えるようになっており、HTML のソースコードを直接編集することなく、教材の作成が可能になっている。

本教材では学習者の認知負荷を低く抑えるため、ブラウザ上で表計算形式でのデータ入力や数式の自然な表現、3D 棒グラフ、2D グラフなどを活用し、これらの要素が順次提示されるようになっている。以下、それらについて解説する。

### ブラウザ上で表計算形式でのデータ入力

本教材では、図1に示すように学習者は表計算にデータを入力するのと全く同じようにデータを入力することができる(キーボードの矢印キーでのセルの移動やコピー&ペーストも可能)。この機能は JavaScript のライブラリである Js spreadsheet([4])により実現されている(JspreadSheet は MIT License であるが、より高機能である有料の Pro 版もある)。



編集モードへ行く

下の表の B2, B3, C2, C3 のセルにデータを入力してください。

	A	B	C	D
1		治った	治らなかった	合計
2	薬剤群			0
3	コントロール			0
4	合計	0	0	0

図1：表計算形式でのデータ入力

### 数式の自然な表現

本教材では、独立性の検定を行う際の計算経過が自然な形で表示されるようになっている(図2を参照)。これにより、学習者は実際にどのような計算が行われるかを実例を通して、理解することが可能である。この機能は、ブラウザ上で TeX のソースをレンダリングする Javascript のライブラリである KaTeX([5]) を用いて実現されている。

### 3D グラフ

本教材では、表に入力されたデータや期待度数の表のデータを 3D の棒グラフで表現することにより、データの視覚的な理解を促すようにしている(図2を参照)。独立性の検定では、入力したデータと期待度数の表のデータとの偏差が特に重要

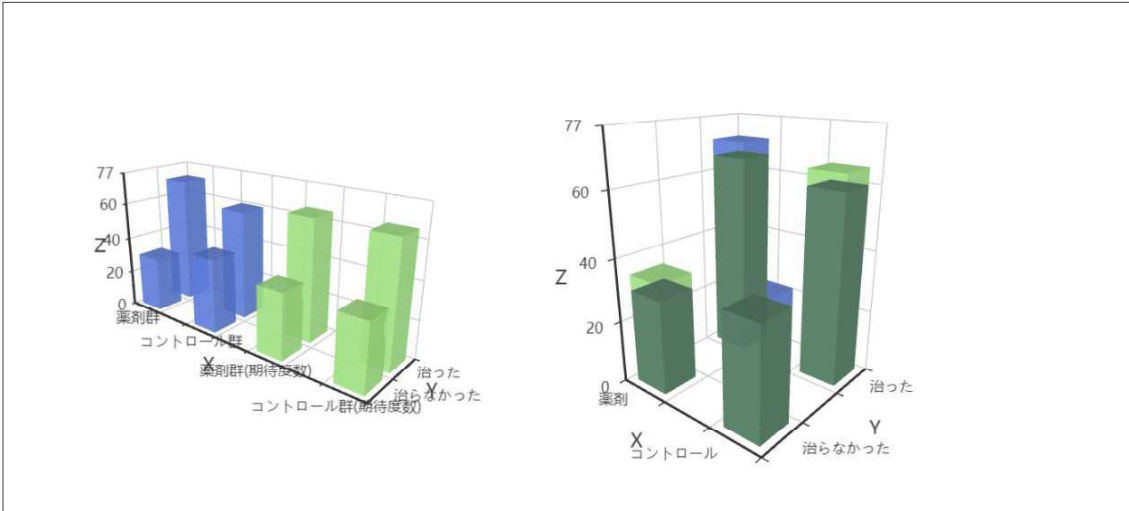
下の表の B2, B3, C2, C3 のセルにデータを入力してください。

	A	B	C	D
1		治った	治らなかった	合計
2	薬剤群	70	30	100
3	コントロール	60	40	100
4	合計	130	70	200

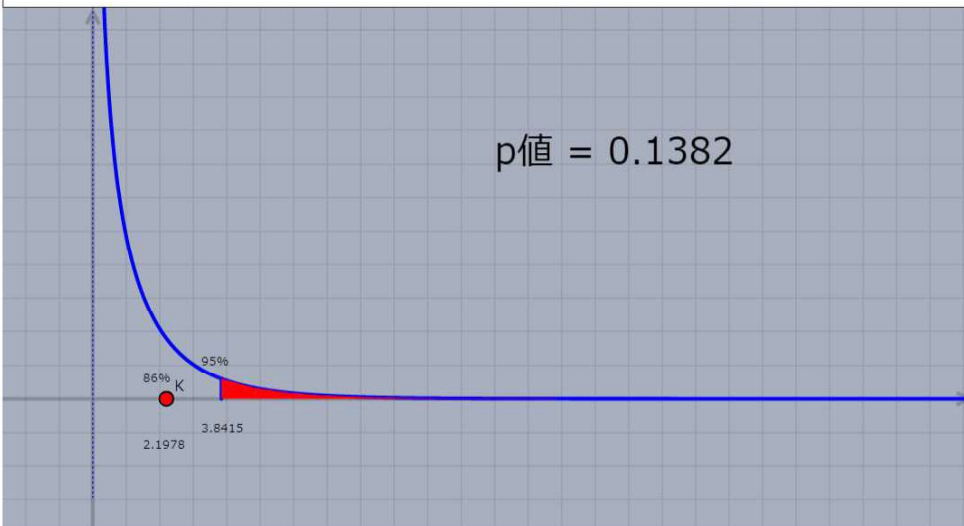
期待度数の表

	A	B	C	D
1		治った	治らなかった	合計
2	薬剤群	65	35	100
3	コントロール	65	35	100
4	合計	130	70	200

グラフ結合



$$\begin{aligned} \chi^2 &= \sum_{i,j} \frac{(a_{ij} - e_{ij})^2}{e_{ij}} \\ &= \frac{(a_{11} - e_{11})^2}{e_{11}} + \frac{(a_{12} - e_{12})^2}{e_{12}} + \frac{(a_{21} - e_{21})^2}{e_{21}} + \frac{(a_{22} - e_{22})^2}{e_{22}} \\ &= \frac{(70 - 65)^2}{65} + \frac{(30 - 35)^2}{35} + \frac{(60 - 65)^2}{65} + \frac{(40 - 35)^2}{35} \\ &= 2.2 \end{aligned}$$



送信

図2：教材の計算実行画面

であることから、これらのデータを3Dグラフで表したものを重ね合わせて表示するようにしている。この機能は、Apache Echarts([6])を用いて実現されている。作成された3Dグラフに対するマウス操作での視点の変更、拡大・縮小が可能である。

## 2D グラフ

本教材では、与えられたデータの独立性判定のための $\chi^2$ 分布のグラフを自動的に作成し、描画している(図2参照)。グラフには $p$ 値や棄却域も描かれており、仮説検定の結果が図から判断できるようになっている。これらの機能は、CindyJS([7])を用いて実現されている。

上記の機能に加え、本教材ではユーザーが表計算のセルに入力したデータをタイムスタンプを付けて記録する機能を持っている。これは入力したデータから、ユーザーのどのような点において認知負荷を抱えているかを判断するためのものであり、現在の所は学習終了時にユーザーが図2の「送信」ボタンをクリックした時にサーバーにそのログデータが送られるようになっている。本論文では、このように取得したログを用いてデータ分析を行う(ユーザーが表計算のどのセルにデータを入力してもその入力データは記録されるが、本論文では、1つのセルのみにデータ入力可能となっている教材を用いている)。

## 3 コンテンツを用いた独立性の検定のフローについて

学習テーマは、薬剤の投与による治癒例の増加効果の有無を確認するため、薬剤を投与した群としていない群(コントロール群)とで治癒例と非治癒例とを計数し、それらをクロス集計表に入力した上で、独立性の検定を行うというものである。本教材では、対象学生の負荷を考慮し、図2の表計算においてB2のセル(左上の色が付いているセル)にのみデータ入力を行えるようにしている(B3, C3を60と40に固定し、C2のセルはB2とC2のセルの和が100となるように自動調整)。つまり、コントロール群における治癒例と非治癒例を60と40に固定し、薬剤群の総数が100という制約条件をつけた上で、治癒例の数値をいろいろと変えて、検定結果がどう動くか観察させるというものである。ユーザーが独立性の検定を行うプロセスは次のようになる。

- (i) 左上の表のB2マスに「薬剤群」における「治った患者数」を数値を入力する。そうすると、D2マスの値が100となるように自動的にC2マスに数値が入力される。また、右上の「期待度数の表」と左の3Dグラフが表示される。
- (ii) 左上の「グラフ結合」のボタンをクリックする。そうすると、左上と右上の表から $\chi^2$ 値が自動的に計算され、計算過程とともに画面に表示される。また、 $\chi^2$ 分布のグラフが $p$ 値とともに表示される。

このように、学習者の認知負荷を抑えるため、データを入力するとほぼ自動的に独立性の検定の結果がその計算過程とともに表示されるようになっている。 $\chi^2$ 分布のグラフ(図2を参照)はCindyJSを用いて表示されており、その際に必要とする $\chi^2$ の値はNumerical Recipes in CのソースコードをCindyJSに移植して計算している。

このようにユーザーが左上の表の B2 マスにデータを入力すれば、そのデータに応じた  $p$  値が自動的に計算されるようになっている。この状態で、データから  $p$  値を計算する通常の問題とは逆向きとなる課題を行う。具体的には、ユーザーにはまず  $p$  値が指定され、ユーザーは計算結果がその  $p$  値に近くなるような B2 マスの値を試行錯誤で探っていく。上で述べたように、このとき入力された B2 マスの値はシステムに保存され、ユーザーが図 2 の「送信」ボタンをクリックするとサーバーに転送される。

## 4 大人数の授業を想定した授業設計と実施状況

先行研究 [1] でも指摘した通り、独立性の検定の全体像を理解する上での難関は、 $\chi^2$  統計量の表示式、特に入力値を変えた際にこれがどのように変動するかという点を理解することであって、ここが教材の改変に当たって最も工夫を要した点である [1]。結果的に、前節に示した通り、クロス集計表や期待度数の表を段階的に表示すると同時に、これらとそのずれを 3D の棒グラフとして提示することにより、 $\chi^2$  統計量の表示式への注意が向けられそうだとの手応えを得ていた。そこで、これらの先行研究 [1] の対象者と同程度の数理解能力を有すると想定される、統計学を履修中の薬学部 2 年生 110 名程度に対して、通常授業の中で本教材を利用することとした。対象者には自身が所有する PC を持参させ、日頃用いている Moodle のサイトにアクセスさせた上で、そこに掲載されているコンテンツを操作させると同時に、理解度を確認するための事後テストを課した。対象者は通常授業の際に Moodle を利用しているが、小テストなどの際に数式入力の実験はしていない一方、CindyJS による動的コンテンツの提示はほぼ毎行われており、その操作も経験している状況であった。

対象者にとって、独立性の検定は新たに学ぶ内容であることから、通常授業で利用している教科書を用いて「独立性」と「相関性」に関する言語的なイメージを確立することが最優先となる。さらにこれを踏まえ、検定統計量として計測値と期待度数との偏差を用いることが自然である点の説明が必要となる。先行研究 [1] においては、これらの説明を与えた後、「帰無仮説」、「対立仮説」、「 $\chi^2$  検定統計量の実現値」、「棄却域」が何に相当するか考えさせる段階を設けたが、本研究の事例は通常授業の一環で、時間的余裕が乏しかったことから、

- (1) 独立性の成立が帰無仮説、相関性の成立が対立仮説に該当すること
- (2)  $\chi^2$  検定統計量の実現値が大きいほど、相関性が強いと考えられること
- (3) 棄却域は、設定された有意水準 (5% を想定) の下で、 $\chi^2$  分布に従う確率分布中、独立性を否定すべき事例を与える領域であること

といったポイントについて、前提情報として教授してしまうフローを採用した。その上で、図 2 のコンテンツの使い方について説明し、10 分間ほど、いろいろな数値を入力して結果を観察する形の探索を行わせた。探索開始前に、事後テストの内容についても予め説明した。これは、薬剤群の総サンプル数を 100 に保ったまま、コントロール群の計測値を探索時の 60・40 から少し変更した際に、 $p$  値が 0.9, 0.1, 0.05, 0.01 の各値に最も

近くなるような、薬剤群の「治癒例」の数値を予想せよというものであった。実際、事後テストで指定したコントロール群の数値は62・38であった。このとき、コントロール群における治癒例が2だけ増加し、非治癒例が逆に2だけ減少するのに合わせて、薬剤群における治癒例を2だけ増加させ、非治癒例を2だけ減少させれば、対応する期待度数も同様の变化を起し、 $\chi^2$  検定統計量の実現値を計算する際の分子部分に相当する偏差が全く影響を受けないことになる。当然、検定統計量の分母にある期待度数も変化はするが、割合として変化は大きくならないことが想定される。以上の点を理解できるかという点が事後テストの主眼であった。与えられた4つの  $p$  値に対し、一貫して治癒例を2つずつ増やせばよいという判断ができる状態であれば、該当の学習者が探索を通して  $\chi^2$  検定統計量の表示式に十分な注意を向けている上に、その振る舞いに関して正しく理解できていると判断される。

コントロール群での治癒例が60であるとする、常識的には薬剤群での治癒例は60以上となるが、数理モデルとして幅広い探索を行わせたいという考え方から、このような「正の相関」を示す事例だけでなく、薬剤群での治癒例が60未満であるような「負の相関」を示す事例についても、探索を妨げなかった。また、薬剤投与群の総サンプル数が100なので、治癒例として100よりも大きな数値を入力することは本来ナンセンスであるが、これも同様の理由から妨げなかった。通常は60から100までの間の整数値が半角で入力されるところだが、全角で数値が入力されたり、数字の0のかわりにアルファベットのOが入力されたりする入力ミスが目立った。これらは、探索経過の可視化にあたり、ログデータを Cinderella に貼り付けた際に、エラーとして明確な警告表示がされることから、比較的修正は容易であった。また、少数ではあるものの、小数値の入力が見られるケースもあった。図3は得られたログデータをもとに、横軸に第1手からの経過時間、縦軸に入力値をとって、探索の推移を可視化したものの一例である。横軸の1メモリは30秒、縦軸の1メモリは10を表している。学習者ごとにログデータを確認すると、操作が第1手のみで終了している例が散見されたが、これはデータの送信トラブルである可能性は極めて低く、周囲の友人の操作を観察していた可能性が高いと考えられる。

さらに、可視化されたグラフにおいて、事後課題として予告した  $p$  値 (0.9から0.01まで) に対応する入力値である60から80までのレンジを3本の水平線で示している。事前に想定されたことではあるが、該当のレンジから大きく外れた数値の入力が多く認められる事例も少なからず存在した。事前の想定としては、探索過程の「ゆらぎ」が学習者にかかる認知負荷の高さを示唆するのではないかということも考えられたが、結果的に、探索の自由度がかなり大きかったことで、ログデータのみによって認知負荷に関する推定を行うことは困難だと判断された。そこで、事後テストの結果に応じてログデータを分類し、正解者と不正解者の間で傾向の違いがないかという観点での分析を実施することとした。

図4は事後テストの問題画面である。この中の表は、探索時に入力される可能性が高い数値 (60から78までの間の整数値) について、結果として得られる  $p$  値をまとめたもので、解答時に参考にすることを許した。コントロール群の数値の変更を度外視して単純にこの表を読んでしまうと、指定された  $p$  値である0.9, 0.1, 0.05, 0.01に最も近く

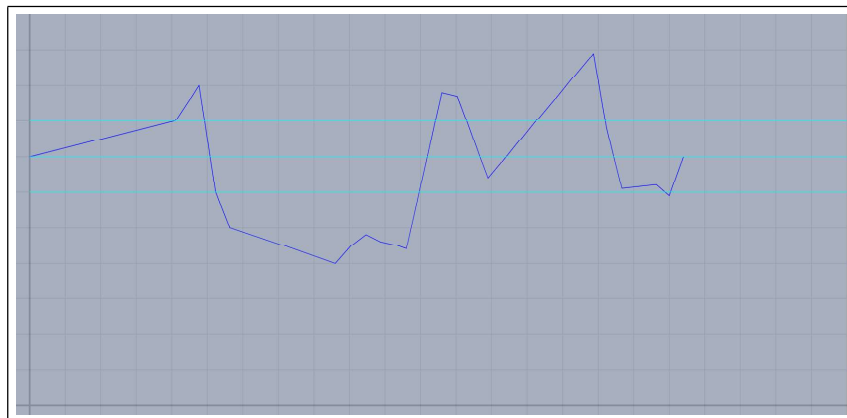


図3：探索経過の可視化

なる治癒例の数値がそれぞれ61, 71, 73, 77となるが、前述の通り、これらを一斉に2だけ増加させた値が事後テストの正解となる。仮に表から読み取った上記の各値をそのまま解答しているケースでは、事後課題の趣旨に関する理解度が疑われるわけだが、その大きな背景として、そもそも探索時に検定統計量の表示式に注意が払われていない状況があると考えられることから、その他の不正解と同等に扱うこととした。

また、図5は解答の入力画面であるが、解答の数値を空欄に入力させた。図4の表を参考資料としたことで、特に制約を加えたわけではないが、60未満の数値が解答されるケースはごく少数にとどまった。一方で、小数値の入力は一定頻度で確認された。

有効な解答109例について、事後テストの解答状況を見てみると、正解率が想定よりもかなり低かったことから、やや緩めに分類することとした。分類基準を含めた結果は表1の通りである。

探索に要した時間を群間で比較したところ、有意差は確認されなかった。先行研究[1]における学習者の状況もふまえると、この結果から、群によらず、コンテンツの中の数式表現や3Dグラフといった要素単体ではなく、これらの要素を統合した理解が目指されたと解釈される。特に、部分正解群については、 $\chi^2$ 検定統計量の数式表現に一定の注意が払われたものの、期待度数の振る舞いについて十分な判断ができなかったものと推測される。さらに、不正解群については、入力値の変更と $\chi^2$ 検定統計量の実現値の動きとの連関については認識できたものの、数式の形に十分な注意が向いていなかったものと想定される。

## 5 操作ログの分析

対象者ごとに図3にあるような探索プロセスの可視化を行ったが、不正解群だけでも操作プロセスは多岐にわたり、理解状況を操作ログのみで特徴づけるのは極めて難しいと判断される。そこで、正解群にはほとんどみられないのに不正解群には一定程度みられるような操作パターンを探してみたところ、正解ゾーン（60から80までの間）を飛び越えて上下の領域を往復するパターンが見出された。該当のパターンでは、単位時間あたりの入力値の変動が大きくなることが想定されるため、全時間帯にわたり、入力値

## クイズ

先ほどの試行では、コントロール群における「治った」例が60例、「治らない」例が40例だった。これがそれぞれ62例、38例で、薬剤群の総サンプル数が先ほどと同じ100であるとき、 $p$  値がそれぞれ 0.9, 0.1, 0.05, 0.01 に最も近づくのは、薬剤群の中の「治った」事例がいくつの場合か、予想せよ。

下の表は、先ほどの試行の結果を一覧にまとめたもので、必要ならこれを利用してよい。

治った	治らない	p値
60	40	1
61	39	0.885
62	38	0.7719
63	37	0.6629
64	36	0.5601
65	35	0.4652
66	34	0.3795
67	33	0.3039
68	32	0.2386
69	31	0.1835
70	30	0.1382
71	29	0.1018
72	28	0.0733
73	27	0.0515
74	26	0.0353
75	25	0.0235
76	24	0.0153
77	23	0.0097
78	22	0.0059

図4：事後テストの問題画面

の総変動量の時間平均 ( $/s$ ) を計算してみることにした。図6は、群ごとに、計算された時間平均の分布をヒストグラムにより表したものである。

想定通り、不正解群の中には、正解群の事例と比べ、単位時間あたりの変動量が大きいものが少なからず確認される。実際、ヒストグラムの横軸に表示された範囲に収まらない事例も複数存在した。ただし、単位時間あたりの変動量が小さければ操作プロセスが良好であるとは必ずしも言えない点に注意が必要である。実際、図7に示される不正解群の事例は、該当の特徴量は小さいものの、ターゲットゾーン（60から80の間）にほとんどヒットしていない。また、図8の事例では、ヒットゾーンの周辺が探索されている上に、該当の特徴量が大きくはないものの、すべての入力の中でゾーン内にヒットしているものが極めて少ないことが確認できる。

このような事例をふまえると、本稿冒頭にふれた、操作プロセスをもとに at risk の学習者を推定するフローを構築する上で、操作プロセスから単一の特徴量を抽出するのは困難で、階層的な診断基準を設定する必要があることが強く示唆される。同時に、教育的観点から探索の自由度を可能な限り担保することと、ログデータに基づいて学習者の探索プロセスを診断できる可能性を担保することとは、トレードオフの関係になりかね



The image shows a screenshot of a post-test answer input screen. It contains four questions, each with a specific p-value and a maximum score of 1.00. Each question has a text input field for the answer and a flag icon for marking the question.

- 問題 1: p 値が 0.9 の場合
- 問題 2: p 値が 0.1 の場合
- 問題 3: p 値が 0.05 の場合
- 問題 4: p 値が 0.01 の場合

図 5：事後テストの解答入力画面

完全正解群	完全正解 または $p=0.01$ の場合のみ不正解のもの	13
部分正解群	全体にフィットは良いが、 $p=0.9$ の場合に 62 よりも小さい値を解答しているもの	8
不正解群	不正解と判断されるもの (このうち、表の値をそのまま読んだものが 35 例)	88
合計		109

表 1：事後テストの結果

ないということも示唆される。

## 6 総合論議と今後の研究の方向性

実験授業前の想定としては、操作プロセスにおけるゆらぎが学習者にかかる認知負荷を反映する可能性があるのではないかという考え方があり、これに基づいて、図 6 にみるように単位時間当たりの入力値の変動量を調査したわけであるが、部分的なものも含めた正解群と不正解群との間で有意差を見出すのは困難で、いささか想定に楽観的過ぎる部分があったと考えられる。しかし、前節に触れた通り、学習テーマの設定から、探索の自由度に制限を加えることにより、有効な教育的介入に向けたより精度の高い知見が得られる可能性は残されている。また、これに伴って、図 6 の調査に用いた measure

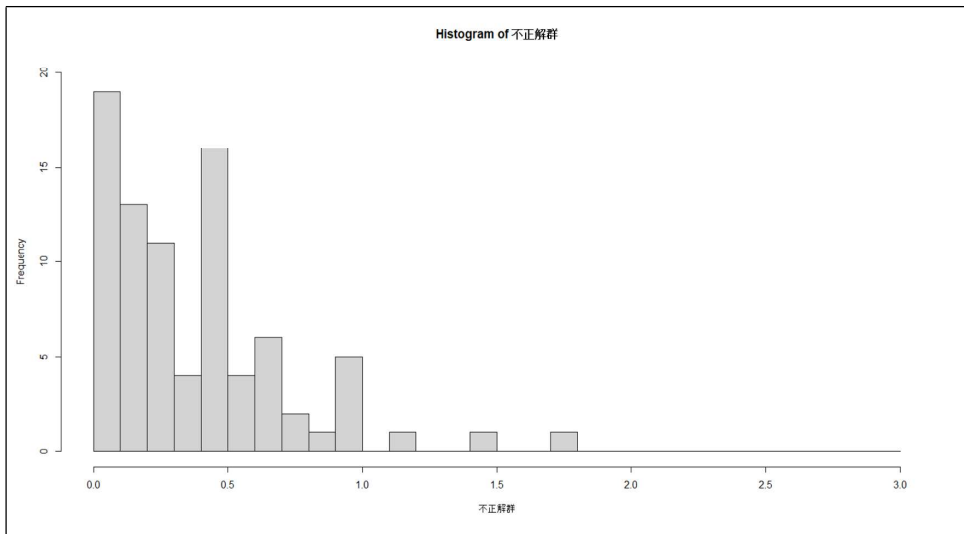
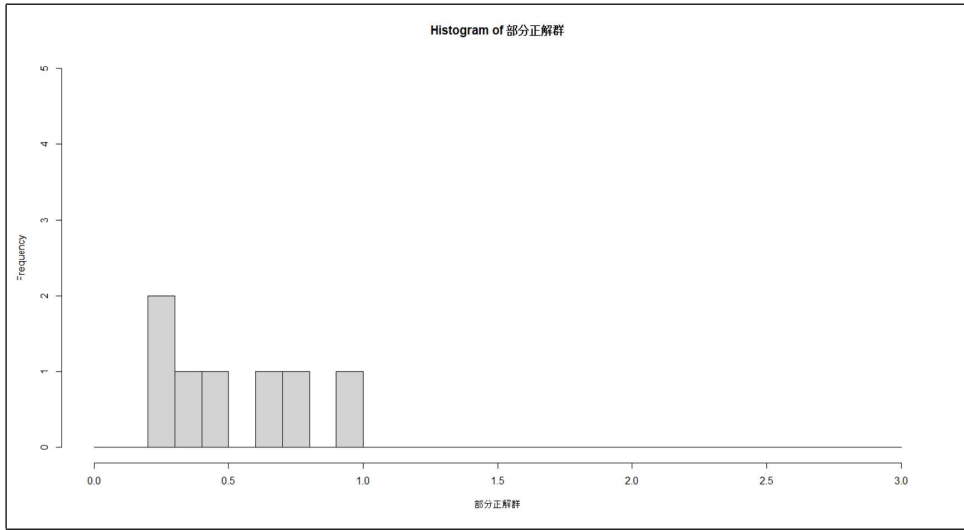
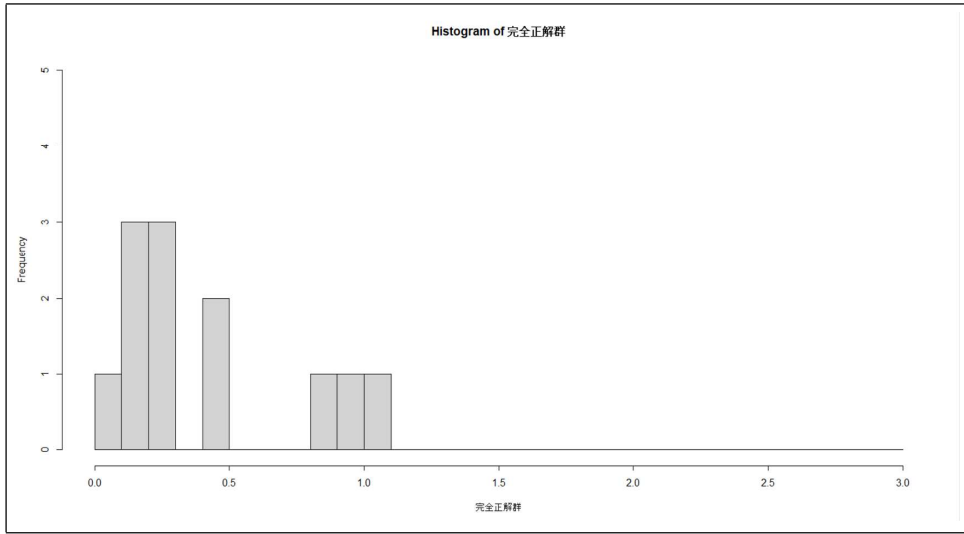


図6：入力値の変動量の時間平均の分布

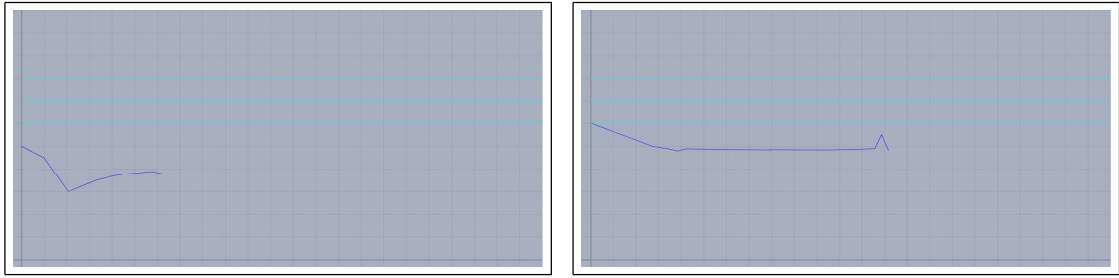


図7：不正解群の操作プロセスの事例（I）

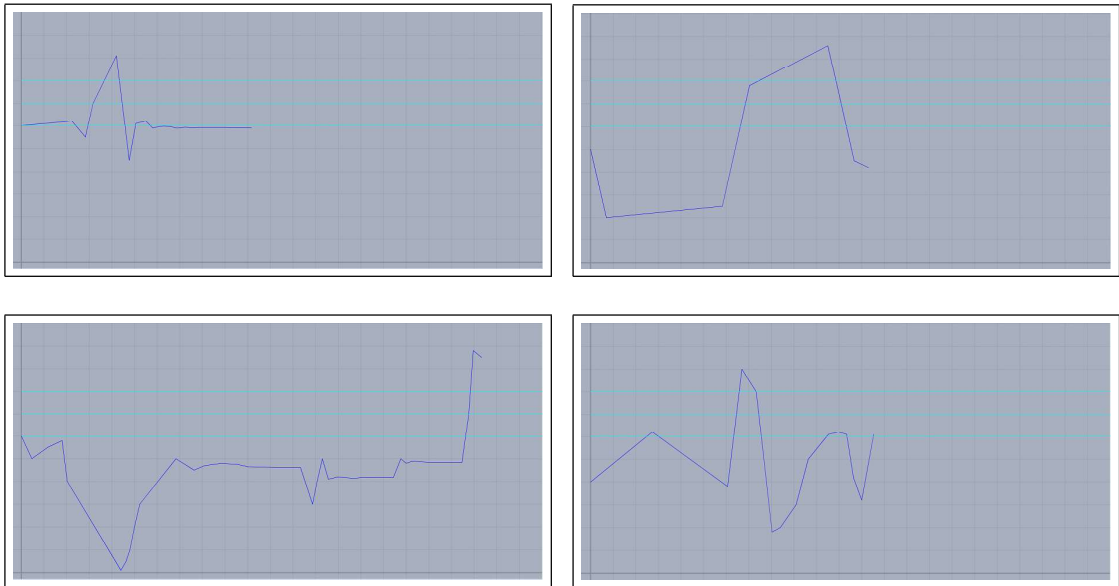


図8：不正解群の操作プロセスの事例（II）

自体を更に変更する必要性が出てくることも想定される。このような示唆に基づき、本稿で紹介したのと同様の環境での通常授業の中で、設定を変更した形でのコンテンツの利用を今後想定している。また、定期試験などの機会をとらえて、授業時に経験した探索が長期的に学習者の思考に及ぼす影響を検証すべく、Delayed Post Test を実施し、その結果と操作時のログデータとの照合を行うことも予定している。

現状では、操作完了後に送信ボタンを押してログデータを送付させているが、技術的には、一定の時間間隔で継続して送信させる方法も可能である。本研究の進展により、at risk の学習者を特定できるような特徴量をログデータから見出すことができる状況になれば、同様の探究活動を行わせる際に、学習支援に向けた新たな方策も可能になると考えられる。即ち、一定の時間的な粒度をもって取得した、授業クラスにおける全学習者のログデータをもとに、上記の特徴量を即時に計算して把握し、問題を抱えた学習者への支援を少しでも早い段階で教授者が行えるようにするというものである。こうした全体像を把握する上で最も可能性が高いのは、web ブラウザ上で全学習者の該当のデータを可視化し、risk の在処を手早く把握できるようにするという方法であり、Learning Analytics の分野では Learning Dashboard と呼ばれて開発が進められているものに相当する [8]。その後の研究の中で、学習データやその分析結果を学習者に直接提示する方法

には問題が伴いかねないとの指摘がなされていることをふまえると [9], まずは教授者を対象に, 学習者の活動ログから得られる情報を可視化するフローを優先的に検討する必要があると考えられる.

## 謝辞

本研究は JSPS 科研費 21K02752 , 22K02951 の助成を受けている.

## 参考文献

- [1] 北本卓也, 金子真隆, 野田健夫 : 独立性の検定に関する HTML をベースとした教材の作成と授業設計, 京都大学数理解析研究所講究録 2236, 掲載予定, 2022.
- [2] J. Sweller, P. Ayres, S. Kalyuga : Cognitive Load Theory, Springer, 2011.
- [3] T. Kitamoto, M. Kaneko, S. Takato : "E-learning system with Computer Algebra based on JavaScript programming language", Proc. of ATCM 2018, Yogyakarta, 2018, 123-133.
- [4] JspreadSheet Official Home Page : URL <https://jspreadsheet.com/> ( 2022 年 9 月 5 日閲覧 )
- [5] KaTeX Official Home Page : URL <https://katex.org/> ( 2022 年 9 月 5 日閲覧 )
- [6] Apache Echarts Official Home Page:URL <https://echarts.apache.org/en/index.html> ( 2022 年 9 月 5 日閲覧 )
- [7] CindyJS Official Home Page : URL <https://cindyjs.org/> ( 2022 年 9 月 5 日閲覧 )
- [8] L. P. Macfadyen, S. Dawson : Mining LMS data to develop an "Early Warning System" for educators, Computers and Education 54(2), pp.588–599, 2010.
- [9] J. P. Tan, E. Koh, C. Jonathan, S. Yang : Learner Dashboards a Double-Edged Sword? Students' Sense-Making of a Collaborative Critical Reading and Learning Analytics Environment for Fostering 21<sup>st</sup>-Century Literacies, Journal of Learning Analytics 4(1), pp.117–140, 2017.