

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

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

1 はじめに

近年、データサイエンス教育の重要性が様々な場面で指摘されている。データサイエンスでは、数理統計学、人工知能 (AI) など、複数の分野を組み合わせ、データ分析を行うが、本稿ではその中で数理統計学に注目する。数理統計学には、統計的推定や検定など様々な応用があるが、特に医療系の学部では、統計的検定が重要 (薬の臨床試験を行った結果、本当に薬が効果があるのかどうか) である。これのためには仮説検定という手法が用いられるが、帰無仮説を設定した上でいくつかの計算ステップを必要とするため、全体像が把握しにくいとの声が学生から寄せられることが多い。そこで、試行錯誤を行いながら仮説検定の手法を理解するための Web 教材を作成することとした。教育現場では、学習者の PC へのアプリケーションのインストールがハードルとなる可能性に鑑み、本教材は HTML をベースとしてブラウザ上で動かすことができる形式とし、パソコンはもとよりタブレットやスマートフォン上でも動作させることが可能な、教材としての汎用性が高いものを目指した。

これまでも、動的幾何システム CindyJS([1]) をベースとしたコンテンツを作成し、シミュレーションを学習者自身で行わせてきた経験はあるが、該当のコンテンツを Moodle 上に実装して利用させる場合、図的なシミュレーションと数式表現による説明とを同期させるにいくといった問題があった。実際、Sweller らによって研究が進められている学習者の認知負荷の理論では、空間的・時間的に統合した形で与えられるべき情報が散在することで学習者の注意が分散されること (split-attention effect) や、視覚と聴覚などの異なる modality にわたって情報が与えられる場合にそれらを統合する必要性が生じること (modality effect)、さらには同時並行で互いに関連付けながら処理されるべき要素数が多くなることで作業記憶領域にかかる負荷が上昇すること (element interactivity effect) などが、学習者の認知負荷に影響する要因として挙げられている [2]。以上をふまえ、本教材では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、動的な可視化などを同期的に利用できるスタイルを採用している。

このように、動的コンテンツに限らない複数の要素を統合した教材を利用するのは筆者にとっても初めての経験であり、実際のところこれが学習者にどのような影響を与え

るかという点については、実際に利用してみてもらわないと予想がつきづらく、教材の改変自体も容易でないという問題があった。近年ヨーロッパを中心に、教員が利用した教材を継続的に収集し、テクノロジーの利用や学習者からのフィードバックが教材作成や授業設計に与える影響を長期的に研究するアプローチ（documentational approach）が出現していることを念頭に [3]，大人数での正規の授業に投入する前段階で、数人の学習者に作成した教材の利用を依頼し、その利用状況を認知負荷の観点から分析した結果に基づいて、教材の改変を試みた。本稿では、こうした改変の状況についてもふれる。

2 HTML をベースとした教材について

本教材では、[4] のシステムを用いて画面の構成やライブラリを活用するプログラムの入力をブラウザ上から行えるようにしており、教材の作成が簡便に行え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 に示すようにブラウザ上で表計算形式でのデータ入力が可能になっている。このため、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([5]) により実現されている。この Js spreadsheet は MIT License であるが、より高機能である有料の Pro 版もある。

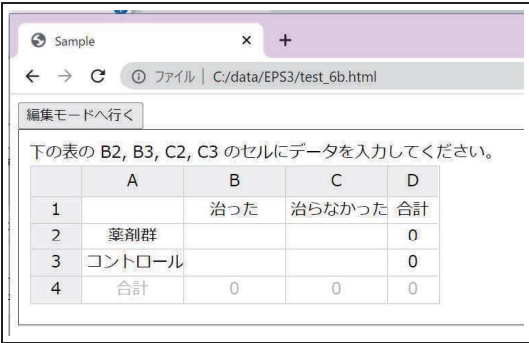


図 1：表計算形式でのデータ入力

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）。学習者はこの計算経過を見ることで、どのような計算が行われるのかを実例で理解することができる。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([6]) を用いて実現されている。

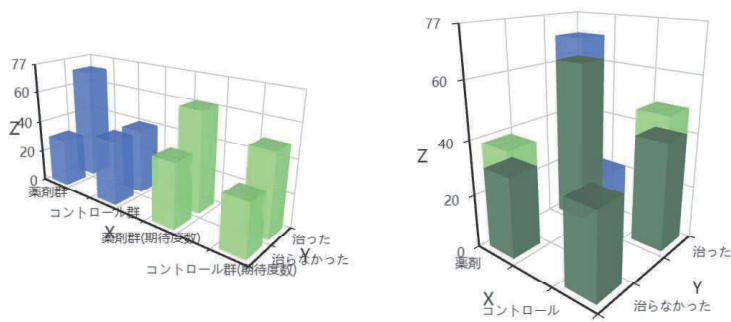
下の表の B2, B3, C2, C3 のセルにデータを入力してください。

	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	40	40	80
4	合計	110	70	180

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	61.1111111	38.8888888	100
3	コントロール	48.8888888	31.1111111	80
4	合計	110	70	180

グラフ結合



$$\begin{aligned}\chi^2 &= \sum_{i,j} \frac{(a_{ij} - e_{ij})^2}{e_{ij}} \\ &= \frac{(a_{11} - e_{11})^2}{e_{11}} + \frac{(a_{12} - e_{12})^2}{e_{12}} + \frac{(a_{21} - e_{21})^2}{e_{21}} + \frac{(a_{22} - e_{22})^2}{e_{22}} \\ &= \frac{(70 - 61.11)^2}{61.11} + \frac{(30 - 38.89)^2}{38.89} + \frac{(40 - 48.89)^2}{48.89} + \frac{(40 - 31.11)^2}{31.11} \\ &= 7.48\end{aligned}$$

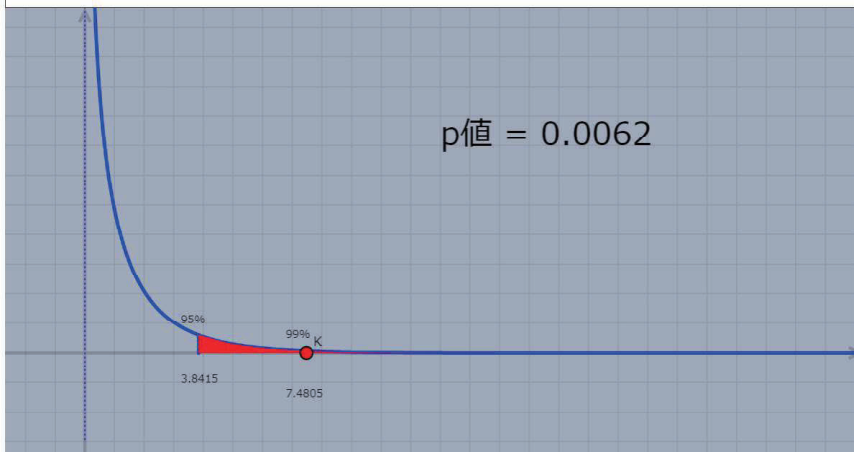


図 2 : 教材の計算実行画面

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で、データの視覚的理解を促すようにしている (図 2 参照)．独立性の検定では、特に入力したデータと期待度数の表のデータとの差が重要なので、これを強調するために入力したデータから作成される 3D グラフと期待度数の表から計算される 3D グラフを重ね合わせたものも表示するようにしている．この機能は Apache Echarts([7]) を用いて実現されている．Apache Echarts はブラウザ上で 3D を含む様々なグラフを描画するための Javascript のライブラリであり、Apache License 2.0 を採用している．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・縮小が可能であり、この機能を用いることで、学習者は 3D 棒グラフを様々な視点から見る事が可能である．

シミュレーション結果の動的な可視化

本教材では、与えられたデータから独立性の検定のための χ^2 分布のグラフを作成し、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、図より仮説検定の結論が図的に判断できる．これらの機能は、CindyJS を用いて実現されている．CindyJS は、動的幾何ソフトウェア Cinderella の機能をブラウザ上で実現させるために開発された Javascript のライブラリであり、これを用いることで Cinderella で作成した図は容易にホームページ (HTML) に変換することができる．

以上のような要素からなる本教材では、薬の効果が本当にあるかどうかのチェックを独立性の検定を用いて行っている．具体的にはその手順は次のようになる．

- (i) 行を「薬剤群」「コントロール群」、列を「治った」「治らなかった」とするクロス集計表を作り、実験結果よりそれぞれのセルにデータを入力する．
- (ii) ステップ (i) の表より、期待度数の表を作成する．
- (iii) ステップ (i),(ii) の表から χ^2 値を計算する．
- (iv) χ^2 値より、 p 値を計算する．また、 χ^2 分布のグラフと有意水準 5% での棄却域を表示する．

ステップ (i) でデータを入力すると、独立性の検定 (上のステップの (ii),(iii),(iv)) の手続きを自動的に実施するようになっている (KaTeX を用いることで計算過程も画面に表示される)．ただし、集計表への入力を行った際に表示されるのは 3D 棒グラフの左側のものまでで、「グラフ結合」ボタンを押下するとそれ以降の部分が表示されるという段階分けをしている． χ^2 分布のグラフも CindyJS を用いて表示するようになっており (図 2 を参照)．その際に必要とする χ^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．また、上のステップ (i) で入力した表に新たなデータを入力し直すと、独立性の検定の計算を自動的にやり直すようになっており、学習者は様々なデータに対して独立性の検定を行うことが可能である．これまでも「独立性の検定」に関する Web 上の教材は存在したが、本教材は学習者が何度もデータを入力して「独立性の検定」の計算をやり直すことができることが特徴である．

3 当初の教材とその改良の経緯

前節に示した教材の前段階として、当初は図3に示すような教材を作成してこれを用いた小規模な実験を何度か繰り返し、その結果を分析した上で、それに基づく教材の改良を繰り返している。改良の中心は、「独立性」「相関性」という概念の言語的なイメージと、前節のステップ (i) と (ii) において作成した表のデータの差から計算される χ^2 値との関連性を学習者にいかに理解してもらうかという点であった。当初は3D棒グラフは用いずに、色の濃淡のみで薬剤群とコントロール群のデータの違いを表現した。学習者が表計算シートへの入力を完了すると、全要素が一気に表示される形になっている。

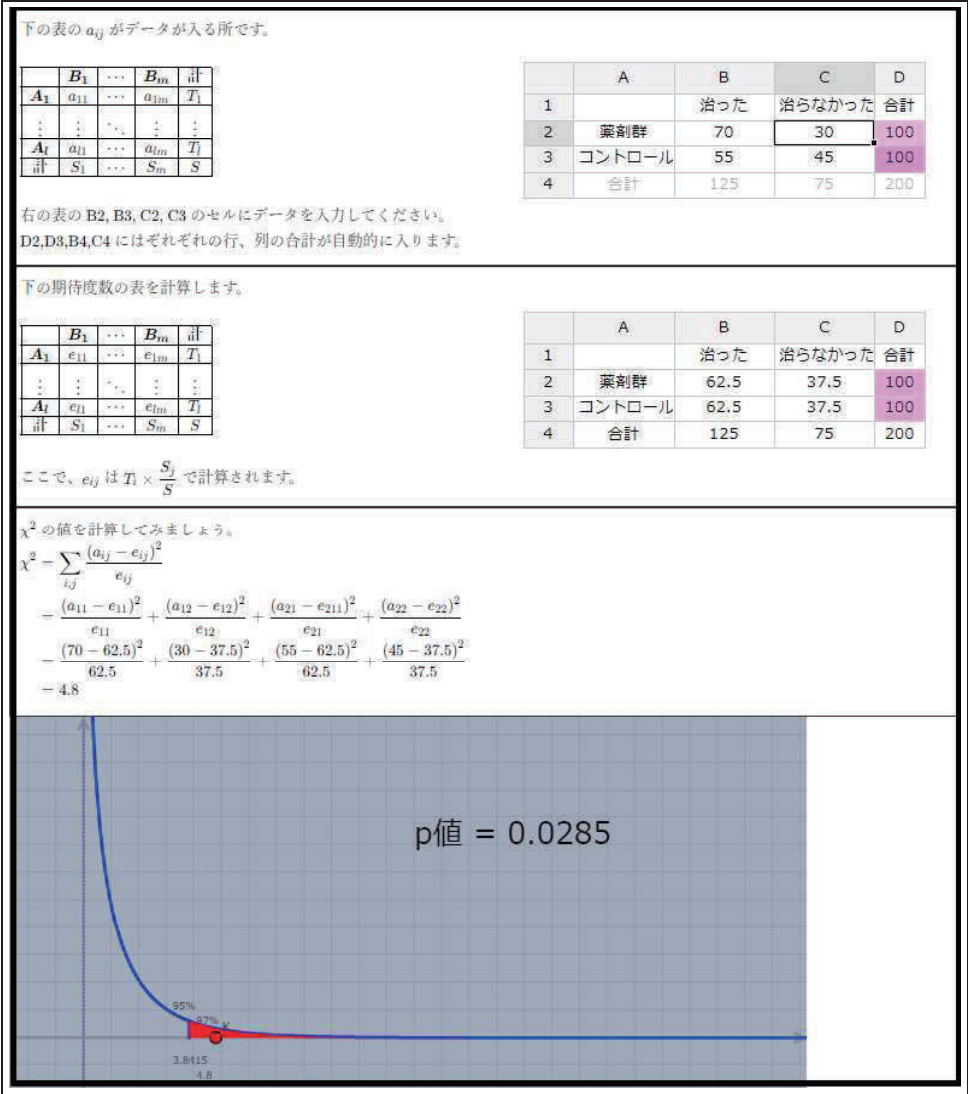


図3：初期段階の教材

この教材の利用状況を確認するため、数理統計を履修済の大学薬学部3年生を対象とした少人数の実験授業を3回実施した。過去の経験上、2名一組での強調学習の形をとることが望ましかったが、新型コロナウイルスの感染拡大が懸念された関係で、初回以外は1名ずつでの実施とせざるを得なかった。また、最終課題の準備を行わせるため、学習者自身がシミュレーションを行えるように、図4に示すコンテンツも併せて準備した。集計表に入力される数値により χ^2 値や p 値がどのように変化するか確認してもらうことを目的としたものだが、入力の手間を省いてより多くの事例にふれてもらえるようにするため、2群のいずれについてもトータルのサンプル数を100とし、コントロール群については効果有・無のいずれのサンプルも50ずつとした。入力した数値と p 値のログについては、コンテンツの中段にある枠の中に表示され、終了後にダウンロードできるようにになっている。

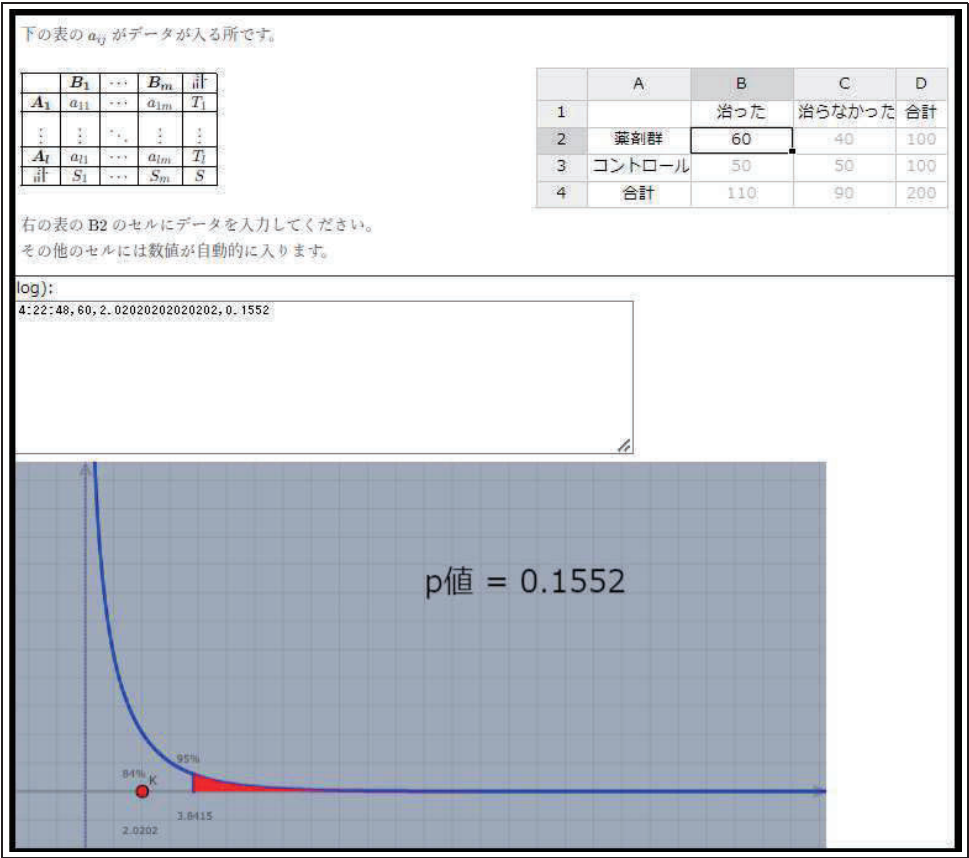


図 4：補助教材

実験授業に際しては、通常授業で用いていた教科書によって検定の手続きや χ^2 分布に関する基本事項を振り返った後、図3のコンテンツを用いて独立性の概念や検定の流れについて説明した。この中では、「独立性」と「相関性」が集計表の行間の比率とどのように関連するかという点を重点的に説明したが、これがある程度了解されたと判断さ

れた時点で、予備課題として図4のコンテンツを用いたシミュレーションを10分程度行わせ、上記の事前説明の内容を自ら確認させた。シミュレーション実施前の時点で最終課題の内容について伝達しておいた。これは、図3のコンテンツの集計表において、コントロール群の数値を55と45、薬剤群のサンプル数を100とした上で、 p 値が0.9、および0.1に最も近くなるような数値を予想させるというものである。シミュレーションの終了後、図3のコンテンツを用いてこの試技を行わせたが、その際、シミュレーションで得られたログの参照を許可し、予想がはずれた場合、どのように修正したらよいかという思考の過程について、できるかぎり詳細な聞き取りを行った。

実験授業に参加した4名の学生は、いずれもCindyJSによる動的なコンテンツを用いた実験授業に過去に参加経験があり、参加者の中でもパフォーマンスが良かった者ばかりである。これは、独立性の検定の概念が決して易しいものではなく、一般的な学力の学生が当初のコンテンツを使いこなすには困難が伴いかねないという判断による。結果的に、3回の実験のいずれについても、入力する数値の変化によって期待度数の表や χ^2 値がどう変化するか、それによって検定結果がどのような影響を受けるかという点について良好な理解が得られたと見られる上に、多くの事例でシミュレートできることでこのような全体像が把握しやすくなったとの回答も寄せられている。これに対し、そもそもこの検定での帰無仮説や棄却域、 p 値の設定について苦労している様子が見えたと他、集計表への色付けや、数式表現のエリア（ χ^2 値の計算プロセス）について殆ど注意が向かなかったとの回答も寄せられた。このような状況からは、まず薬剤投与の有無と疾病治癒との相関が強まると2つの要素の独立性が弱まる点の言語的な理解が先行し、そのような状況下でクロス集計表と期待度数の表との乖離が広がり、 χ^2 値が増大していくメカニズムの直観的理解が続いたと判断される。一方で、こうした理解を、独立性の成立を帰無仮説、 χ^2 値を検定統計量とする仮説検定の枠組に統合する段階で相当大きな認知負荷がかかったことをうかがわせる。実際、3回の実験とも、予備課題の完了までに予想を上回る時間を要し、コントロール群の数値の変化に対して、与えられた p 値を実現するための入力値をどのように調整したらよいかという判断に手間取ったことを示している。以上の結果は、定性的に独立性・相関性と p 値の動きとの連関について一定の理解はあった一方で、検定統計量の関与についての注意がやや欠落しており、コンテンツに提示されたいくつかの要素の中でsplit-attention effectによって数式表現エリアに十分な注意が向かなかったことの影響が推察される。さらに、特に1回目の実験の際には、学習者の行動の面でも顕著な特徴が確認された。図5は、最終課題に取り組んだ際に、薬剤投与の有無との相関性が高まると χ^2 値が増大していく様子を理解する上でジェスチャーを援用している状況を表す。この結果は、独立性・相関性と χ^2 値の動きとの連関に関する直観的理解を検定統計量の計算メカニズムの理解に統合する上での困難を、視覚や聴覚とは異なるジェスチャーというmodalityを用いて補完しようとしていた実態を表すものと理解される。

以上のように、数理的能力に比較的秀でていると想定される学習者をもってしても、図4のコンテンツを用いて独立性の検定のメカニズムの全体像を把握する際に困難を伴う実態に鑑みると、コンテンツとそれに伴う授業デザインにさらなる工夫が求められることが想定された。

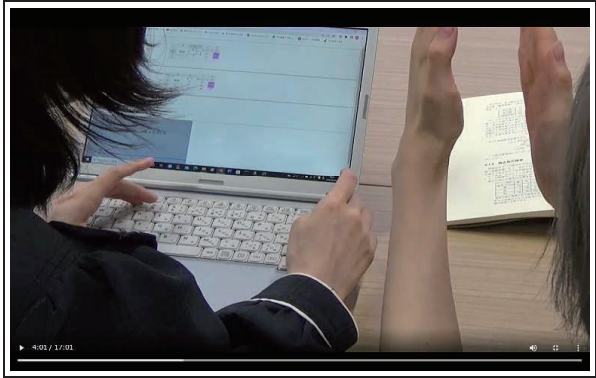


図 5：最終課題に取り組んだ際のジェスチャー

工夫を要する点の第一は、全要素を一度に提示することで split-attention effect が働くことへの対策であり、段階的な提示によって対応することが想定された。さらに、検定統計量の計算プロセスとの結びつきを理解させるためには、クロス集計表と期待度数の分布を可視化することが有効であると考えられた。そこで、図 6 のように、3D の棒グラフを用いることによって 2 群のデータの違いを表現することを試みた。

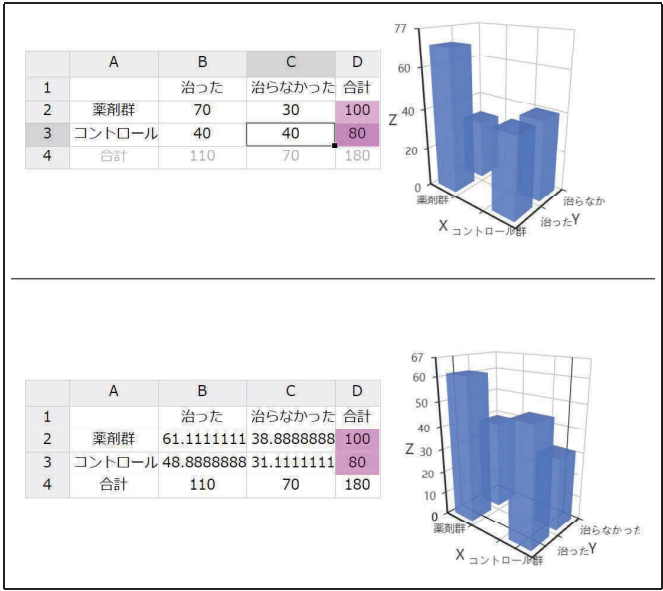


図 6：3D グラフによる 2 群の差異の可視化

このような工夫によって「独立性」と「相関性」の対比を視覚的にとらえやすくなると期待される一方、1つの 3D グラフの視点を変更した際に、もう 1つの 3D グラフも連携して視点を変更できた方が良いのではないかという見方が浮上した。このため、図 7 のように 3つの 3D グラフを横に並べて表示することにした。

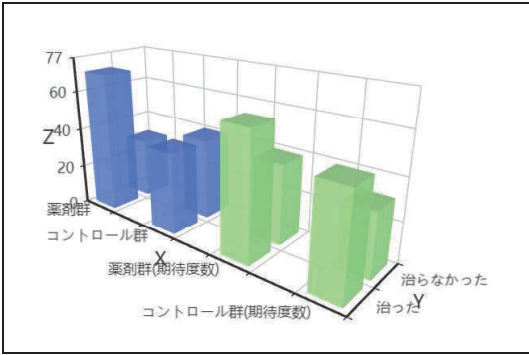


図 7：2つの分布の違いに関する並列的な可視化

以上の工夫によって、「独立性」と「相関性」の視覚的理解が促進されることが期待される一方で、 χ^2 値の計算プロセスとの結びつきを意識させる上ではさらなる改善が必要だとの見方が出た．そこで、図 7 の 2 つのグラフを重ね合わせた物を作り、2 つのグラフのデータの差が重要であることに気づかせたらどうかということになった．結果として作成されたのが図 2 に示した教材である．

4 改定版による実施事例と認知負荷の観点からの分析

本節では、図 2 に示した改定版の教材を用いた実験授業の状況を提示し、その結果を認知負荷の観点から分析することを試みる．認知負荷の推定については、主要なタスクと並行して secondary なタスクを課し、それに対する反応時間を計測したり、視線の動きや心拍などの physiological なデータを計測するなどの方法が提案されつつあるが、詳細な推定が可能となる状況とは程遠いのが実情で、ヒアリングに基づく subjective な感覚の評価や、主要なタスクそのものに対する解答時間のデータを用いた古典的方法が依然として主流である [8]．以上を踏まえ、本教材の利用状況に基づいた学習者の認知負荷の分析にあたっては、対象者へのヒアリングの結果とタスクの実行時間のみを考慮することとした．対象者は、前述の実験授業と同様に数理統計を履修済の大学薬学部 3 年生 2 名に加え、別の学部で数理統計を履修した学習者 1 名の併せて 3 名であり、平均的な数学の学力をもつと想定される学生であった．ただし、最後の 1 名については、新型コロナウイルス感染拡大の影響で、数理統計の履修が完全オンラインの環境で行われ、具体例をもとにした学習の動機付けに時間が割かれた反面、数値例による検定の手続きの経験が少なかったとの申告があった．

このような学習歴の違いは、事前説明段階での理解に大きな違いを生んだ．実際、前者の 2 名については、前節の実験授業と同様に、帰無仮説・棄却域・ p 値がこのケースで何に相当するかという点の理解に厳しいものがあったのに対し、後者の 1 名については、事前説明段階からクリアに理解されていたものと推察された．その一方で、3 名に共通した傾向として、前節の実験授業と同様に「独立性」「相関性」と χ^2 値の動きとの連関について、言語的な理解が先行したものの、前節の実験授業と比べ、検定統計量の計算

プロセスがより明確に意識されていたことが観察された。ヒアリングの結果、この違いは、3D グラフに表示される「残差」と χ^2 値の計算式に現れる数値とがクロス集計表への入力の変更に連動して変動することの観察を通してもたらされたことが確認されている。また、いったん上記の連関性が理解された後は、 χ^2 値の計算プロセスは意識されなくなったことも確認された。このことは、3D の棒グラフでの可視化によって、クロス集計表の行間の比率の差異と χ^2 値との連関の理解が「カプセル化」され、その後は「独立性」「相関性」の言語的イメージと χ^2 値の動きとが直接的にリンクされたことを強く示唆する。その背後では、3D の棒グラフを段階的に提示することで、作業記憶領域で並行して処理する要素の数を減らし、認知負荷の過度な上昇を抑制するという element interactivity effect が働いたものと考えられる。こうした効果は、図4のコンテンツを用いた予備課題に要した時間に鮮明に反映されたと考えられる。実際、当初版を用いた3例での所要時間がそれぞれ13分30秒、8分10秒、6分であったのに対し、改定版を用いた3例ではそれぞれ9分45秒、7分30秒、5分45秒となり、全体的に短縮されている。もちろん、この時間がすべて探索に用いられたわけではなく、完全な比較は困難であるが、前者の3名が後者の3名と比べて数理的能力に関して優勢であると想定される上、最終課題の所要時間に関しても、前者がそれぞれ6分20秒、7分、5分であるのに対し、後者のうちの最初の2例でそれぞれ4分30秒、6分40秒と全体的に短縮されている結果とも符合し、コンテンツの改定によって学習者の認知負荷を下げる効果があったことが示唆される。

ところが、予備課題での所要時間が最短の5分45秒であった最後の事例について、最終課題の実行時に予想外の事態が発生した。 p 値が0.9に近いケースが比較的早い時間に実現された一方、その後 p 値が0.1に近いケースを実現するまでにかなりの長時間を要し、トータルの所要時間が9分40秒となった。図8はその際の様子であるが、左図にある p 値が0.1489の状況を実現した後、本来は p 値を減らす方向に入力を変更すべきところ、右図にある p 値が0.2501の状況に遷移してしまっており、探索が迷走していた状況をうかがわせている。

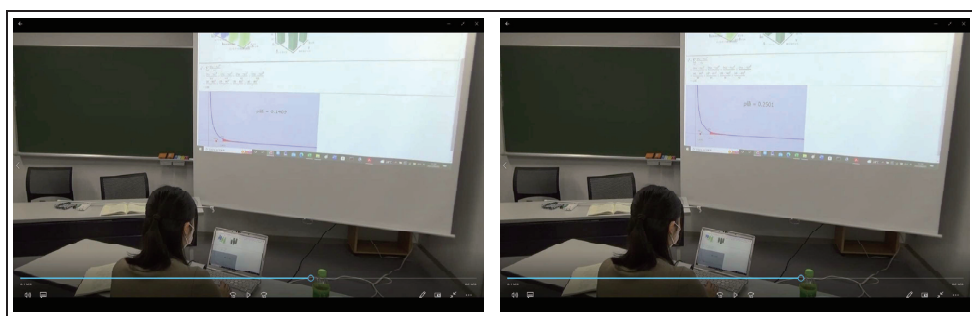


図8：最終課題実施時の状況

該当の学生への事後のヒアリングの結果から判断すると、予備課題実施当初は、クロス集計表への入力の変更と p 値の変化の連関について一定の理解ができていたものの、 χ^2 統計量の計算プロセスに関する把握が十分でなく、最終課題実施の段階になると上記の

連関に関する理解が失われてしまった状況が見受けられた。前述した通り、該当の学生については検定に関する計算手続きの経験が不足していたことをふまえると、クロス集計表への入力値と χ^2 値とが連関して「動くこと」は認識できていたものの、 χ^2 の計算プロセスに現れる偏差の増減が影響するメカニズムまでは十分に把握できていなかった可能性が疑われる。他の学生と比べて該当の学生が、事前説明の段階で帰無仮説や棄却域に関する確かな言語的理解を有していたことと考え併せると、科目履修時点において数値例による検定の手続きに関する経験が不足したという学習歴が影響したことは想像に難くない。

本稿で紹介してきた事例は、正規の授業で教材を活用する手前の予備的な試行であって、実施事例が多くないのは確かだが、一口に「テクノロジーを活用した探索によって仮説検定を理解した」と言っても、その内容はかなりの重層性をもつことを強く示唆するものである。すなわち、通常の授業において検定統計量の理論的位置づけや実現値の計算に重点が置かれ、学習者にニーズに即した検定のイメージ付けが不足すると、新たな種類の検定に出会った際に、その手続きの言語的イメージの獲得に支障が生じかねない。他方で、検定の実用的意味の説明が先行し、検定統計量の理論的な理解が不足すると、実例を用いた検定の実施の際にその全体像が把握されないというリスクが生じる。近年のデータサイエンス教育では、どちらかという統計の実用的意義が強調され、その数理的背景が置き去りにされがちであることが否定できない。できるだけ多くの学習者をデータサイエンスにふれさせたいという必要性からやむを得ない面はあるのだが、この状況でテクノロジーを用いた数理統計的な探索を行わせた場合、学習者の学習歴や数理的な能力によって、困難をかかえる部分に相当大きなバラエティが生じる可能性を想定した上で、教授者側もそれに応じた教育的介入を行っていく必要があるということ、本研究の事例は示している。

5 今後の研究の方向性

改定版のコンテンツを用いて平均的な学力の学生での実験授業がなんとか完了できたことから、実用段階にかなり近づいていることは確かだと考えられるが、実際にこれを100人以上の規模の授業で利用すると、さらに検討すべき課題は少なくない。現段階では、限られた時間の中での複数のコンテンツを用いた授業進行が困難であることから、事前説明と学習者による予備的な探索を図2のコンテンツに集約させる形での授業設計を想定している。また、探索のプロセスを把握して適切な支援に生かすため、操作ログを蓄積して送信させた上で、事後テストを課し、学習者個々の理解度を把握することを計画している。

本研究の実施事例からは、特に仮説検定のような複雑な学習課題についてテクノロジーを用いた探究的な学習を行わせる場合、大人数の通常授業に投入する前段階として、少人数形式の授業に試行的に投入し、学習者の探索状況について詳細に分析しておく必要があることが示唆される。これについて、協調学習を行った事例が示す通り、学習者の aptitude によって、認知負荷の上昇を抑える上で活用される modality に多様性がある点に注意が必要である。該当の事例では、2人の学生のいずれもが、「独立性」や「相

関性」の言語的イメージと検定統計量の動きとの関係を理解する際に、ジェスチャーを多用して認知負荷の上昇を抑えている様子がたびたび観察されている。これは、この実験とは別の機会に CindyJS のコンテンツを用いた協調学習を行わせた際にも顕著に確認された傾向で、実際、近年の研究でも、学習活動においてジェスチャーの果たす役割の重要性が指摘されてきている [9]。このようなタイプの学習者の思考過程を確認する上では、単にその発話内容だけでなく、探究プロセスの中で発生したジェスチャーの関与まで調査する必要がある。もとより、録画によって質的な調査をすることは必要であるが、操作ログと同様の時間的な粒度をもって、手指の位置情報を取得できれば分析上の大きな助けになることが想定される。このため現在、図9に示す通り、深度センサー付きカメラを用いて図10に示す画像を録画し、ここからタイムスタンプを付けて手指の位置情報を取得する方策について検討している。

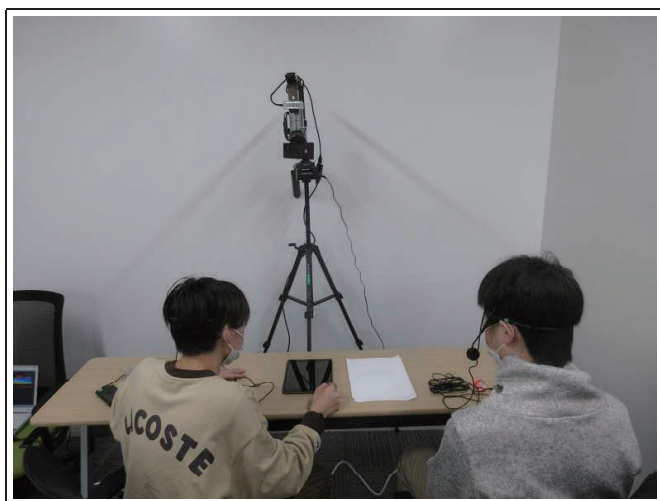


図9：深度センサー付きカメラを用いたジェスチャーのログの取得

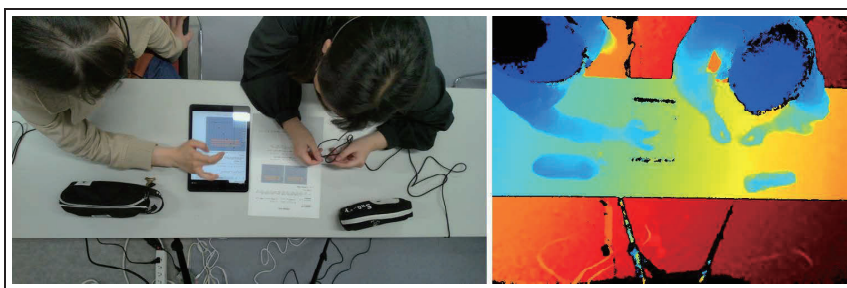


図10：得られる録画画像

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] CindyJS Official Home Page : URL <https://cindyjs.org/> (2022 年 9 月 5 日閲覧)
- [2] J. Sweller, P. Ayres, S. Kalyuga : Cognitive Load Theory, Springer, 2011.
- [3] G. Gueudet, B. Pepin, L. Trouche : From Text to Lived Resources: Mathematics Curriculum Materials and Teacher Development, Springer, 2013.
- [4] 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.
- [5] JspradSheet Official Home Page : URL <https://jspradsheet.com/> (2022 年 9 月 5 日閲覧)
- [6] KaTeX Official Home Page : URL <https://katex.org/> (2022 年 9 月 5 日閲覧)
- [7] Apache Echarts Official Home Page:URL <https://echarts.apache.org/en/index.html> (2022 年 9 月 5 日閲覧)
- [8] R. Zheng : Cognitive Load Measurement and Application: A Theoretical Framework for Meaningful Research and Practice, Routledge, 2018.
- [9] J. Nathan, C. Walkington, M. Alibali et al. : Actions speak louder with words: The roles of action and pedagogical language for grounding mathematical proof, Learning and Instruction 33, 2014, 182-193.