

ユビキタス分析環境が拓く「文系からの文理融合型人材育成」の可能性

天野 徹¹

Email: amano @meisei-u.ac.jp

*1: 明星大学人文学部人間社会学科

◎Key Words

量的社会調査実習, アジャイル型仮説検証, ユビキタス分析環境, 数学が苦手な文系学生

1. はじめに

IoTの進展やビッグデータの蓄積などにより、我々を取り巻く情報環境が激変する中、データサイエンティストの要請が社会的な課題となっている。やLLMの飛躍的な発展に代表される、情報処理の精度とスピードの向上により、文系型のDX人材すなわち、文系的センスのもとにデータの分析・解釈やAIの活用ができる人材の育成が求められるようになった。そしてSociety5.0は、「個人の問題関心・学習の進度に応じた教育」を容易にする、情報環境を提供しつつある。こうした状況下で、文理融合型人材の育成に向けた動きが進んでいる。ただし、文系学生の多くは高校在学中の段階ですでに、「数学に対する苦手意識」を持っていることから、大学における文理融合型人材の育成は大きな困難に直面しているものと思われる。

そして、標準化調査によって得られた数値データの分析を行う授業においては、多くの学生はさらなる困難に直面することになる。データの分析を行うには、分析に必要なソフトウェアがインストールされている大学構内の特定の施設に滞在しなければならず、時間的・空間的に制約される。統計モデルによる検定作業を行うには、複雑な操作すなわち、表示されるメニューの中からの確かな選択肢を間違いなく選んでいく必要があるが、表示される文言は文系学生が難解と感じるものが多い。分析を進める上で新しい変数の合成を行おうとすれば、新たな操作を覚える必要が発生する。このように、調査データの分析・評価・解釈とは全く関係のない要素が、文系学生のデータ分析作業の障壁だった。これもまた、文系型のデータサイエンティストの育成の、大きな阻害要因といえよう。

本稿では、このうちの後者、すなわちデータ分析環境に起因する人材育成上の困難を克服するために「ユビキタス分析環境」を提供する試みが、学生自身によるアジャイル型仮説検証に対する意欲を引き出し、文系型データサイエンティストすなわち、文系からの文理融合型人材の育成について、どのような意味で貢献したかについて、考察することにした。なお、数学に苦手意識を持っている文系の学生たちが、統計モデルと検定理論を用いてデータの分析を行うにあたっては、文系学生に相応しい統計モデルおよび統計的帰無仮説の理解と、検定理論を用いた仮説検証を行う上で必要な帰無仮説・対立仮説の言い回し、検定結果として導き出されるp値の評価についての適切な表現法の使用が不可欠である。小生の授業では、学生たちがこれらのことについて「自分の頭で納得して使う」ことができるようになるよう、文系の世界観に合わせた形で確率・統計の理論を再編集しての解説を行なっているが、本稿での言及は、必要最小限の紹介に止めたい。

2. 文系学生向けの確率・統計教育のリデザイン

2.1 教育コンテンツのリデザインについて

数学嫌いの学生が生まれる理由にはいくつかの理由があるが、社会調査データの分析を行う学生には、単にソフトウェアの使い方とアウトプットの評価法をテクニックとして学ぶのではなく、分析に用いる統計モデルに関係する理論を、最も初歩的なレベルから理解し、スキルとして活用できるようになってもらうことが望ましい。ただし、数学は「積み重ね」の学問であるから、従来の方法面に固執すれば、中学レベルの確率統計から学び直す必要がある。そして、中学高校での数学の授業が、文系学生の数学嫌いの大きな原因となっていることを考慮すれば、大学で数学嫌いを克服するためには、教育内容と方法を、数学嫌いの文系学生にむき再構築する必要がある。しかしながら、社会調査実習の講義には社会調査協会から要請された縛りがあり、数多い統計モデル全てについてそうしたことを行う時間を確保することは不可能であった。

そこで、小生の担当する授業では、学生に使用させる統計モデルを、ほぼ全ての変数について適用可能なカイ二乗検定と、クロス表の各セルにおける出現度数と期待度数の乖離の程度を評価するハーパーマンの残差検定に限定し、この二つの統計モデルの理解に必要な確率統計理論を解説した上で、各統計モデルがどのような理論に基づいて作られているかを理解させるという方法をとった。¹⁾

カイ二乗検定を理解するには、二項分布と乗法定理・加法定理及び、正規分布による二項分布の近似について学ぶ必要があり、当然のことながら正規分布を理解する必要がある。近似の妥当性を理解させるには、パスカルの三角形や二項定理を用いて視覚的に理解することが必要であり、カイ二乗分布を理解するには、Z得点の二乗和の出現確率分布がセルの数によって変化する理由を理解することが必要だ。ただし、カイ二乗値の計算式はそもそもカイ二乗分布を導く数式と事なっており、自由度により数学的操作を行うことにより近似を行なっていることを、文系の学生に理解してもらうことは難しい。しかし残差検定の理論を用いれば、カイ二乗検定のモデルの理解は容易になる。残差検定は有意差が認められた場合のクロス集計表の分析に役立つことから、こうした解説方法は分析結果の解釈を行わせる上でも、大変有効であった。

但し、これは「調査データの分析技法を本質的に理解する」ことを目的とした、確率・統計の教育体系の再編に過ぎない。文系学生の数学嫌いの克服のためには、確率統計の理論の展開を、その背景となる社会思想と関連させながら解説する教育を行うべきで、テキスト・著書²⁾を作成して行なった教育実践では、高い効果が認められている。

2.2 ユビキタス・データ分析環境の提供について

昨今における学生の多様化は、社会学を学ぶ学生たちの問題関心の多様化と、自らの問題関心にそぐわない課題を課されれば簡単に授業放棄するという現象をもたらしている。一方で、VUCA ワールドの進展は、一般化された設問による調査と総花的な調査報告書が、ソリューションに繋がらないという事態を生み出している。このような事実をふまえれば、社会調査実習の授業は、学生が自らの問題関心に基づき、自由にテーマを設定し、自らの分析結果を深掘りするプロセスを支援する形に必要がある。そして、学生たちが自らテーマを決定し、アジャイル型の仮説検証作業を行い、特に重要と感じた事柄について自由に深掘りができるようにするためには、時間や空間そして費用の制約のない環境、つまり、いつでもどこでも、難解で面倒な操作を覚えることなく、使い慣れた機材で分析を繰り返せる環境が必要となる。

さらに、こうしたデータ分析環境は、「SPSS for Windows」のように、多層化されたメニューから難解な専門用語のラベルを次々と選んでいかなければ分析ができないようなものではなく、簡単な操作で気軽に仮説検証を繰り返せる類のものであることが望ましい。

なぜならば、(1)文系学生にとって、数理モデルの検定に用いる様々な手法の全てについて、概念と名称に慣れることは困難であり、(2)大学の施設の特定の機材から限られた時間・限られた人数しか同時利用できないソフトに慣れることは不可能なことに加え、(3)GUI によって見た目には使いやすくなっているにも関わらず、SPSS など汎用分析ソフトのメニューには幾重もの階層があり、その中から分析目的に応じた機能を間違いなく選択していくには慣れが必要であること。そして、(4)新しい変数の作成が必要となれば、複雑な操作あるいは簡単なコードの作成が必要となり、それが「より深い分析」の障害となるからである。そして重要なことは、こうした阻害要因が、社会学理論および文系のセンスを活かした社会調査データの分析能力の習得という、講義の本来の目的とは全く関係がないということだ。

文系型データサイエンス人材とは、データサイエンスのモデルに合わせて社会現象を解釈する人材ではなく、文系的センスに基づき社会現象を理解するための多様な仮説を数多く生み出し、その検証のための道具としてデータサイエンスを活用できる人材であるはずだ。今後、AI の発達等によりデータ分析者自らがメニューを選択していくという行為自体が不要になっていく近未来を想定すれば、今現在の段階でこそ、こうした不毛なスキル習得をスキップして、量的データ分析能力を身につけた文系学生を育てるための活動が、広く行われるべきであろう。

幸いなことに、python についてはクロス集計用ライブラリ、カイ二乗検定を行うライブラリ、残差検定を行うライブラリが公開されていた。また、商用サービスを活用すれば、python で作成したプログラムを用いてデータ分析の講義専用の web サイトを構築できる状態であった。いつでもどこでも PC やスマホでデータ分析ができる環境が提供されれば、自由にアジャイルな仮説検証を行うことを通して、VUCA 時代に対応できる文系型データサイエンス人材へと育つ者が現れる。そうした仮定に基づき、一ヶ月の期間を費やして、web サイトの構築を行なった。

3. データ分析プロセスの支援について

3.1 データ分析サイトの機能およびUIについて

量的手法を用いて社会調査を行う場合、自由回答の設問を除けば、変数の種類は(a)離散変数と(b)連続変数の二つであり、測定の尺度は(a-1)名義尺度、(a-2)順位尺度、(b-1)間隔尺度、(b-2)比率尺度に分けることができる。データ分析の入門者が視覚的に最も理解しやすいのは、連続変数の間の関係を調べる数理モデルである回帰分析・相関分析であろう。統計的分析手法を解説した書籍には、X 軸と Y 軸に二つの変数を割り振り(回帰分析の場合は独立変数と従属変数の区別をつけるが、相関分析の場合は両者の区別がない)、観測値を用いて描いた散布図と相関係数の値の組み合わせについて、いくつかのパターンを用いて関連の方向性と強弱の評価方法を示しているものが散見される。しかしながら、実際の社会調査データでは、離散変数を用いた分析を行う必要があるケースが少なくない。また、連続変数は離散変数に変換して分析を行うことも可能であり、その意味で離散変数を用いた分析モデルの方が適用範囲は広い。そして、単純クロス集計では、疑似相関などによって判断を誤る可能性があることから、ダブルクロス集計ができる環境が必要となる。

これらのことを考慮して、小生が担当する授業では(1)単純クロス集計とダブルクロス集計ができる、(2)出力されたクロス集計表一つ一つについてカイ二乗検定ができる、(3)クロス集計表の一つ一つのセルについて「行%の値」の出力ができる、(4)クロス集計表の一つ一つのセルについて、「残差検定の値」の出力ができる機能を備えた集計環境が必要となる。そこで、python を用い、公開されているライブラリを活用してプログラムを作成して、web サービスの形で学生に使用させることとした。

3.2 合成変数の作成に関わる学生サポートについて

ただし、こうして構築したプログラムは、連続変数を離散変数に変換する機能や、複数の変数の値を元に新たな変数を作成する機能を持ち合わせていない。したがって、調査によって得られたデータのうち、離散変数同士の関係についての仮説検定しか行えないという欠陥があった。

この授業の主題は、学生がいつでもどこでも簡単な操作で活用できるデータ分析環境を提供し、受講生一人一人がアジャイルな仮説検証を行えるようにすることで、文系的センスを活かせるデータサイエンティストを育成すること。合成変数の作成のために必要なアプリケーションの操作はそれを阻害する恐れがあるので、取り除かなければならない。幸い、データ分析環境を提供するために構築した web サーバーのディレクトリ構造やファイル形式については十分な情報が得られている。したがって、新しい変数の作成のためのアルゴリズムさえあれば、新変数の作成と分析データの換装は比較的容易であった。

授業では、受講生から新変数の合成の要望が出された際に、新変数を作成するためのアルゴリズムを提案させるが、変数の作成とデータファイルの換装は教員が行う形とした。新変数の作成のアルゴリズムは LMS 上に提出させ、授業時間中に新変数を作成・データを換装する。換装が終われば学生たちは即座に集計作業を行い、変数の効果を確認した。学生達は次第に、想定された結果が得られなければ、即座に代案を考えて提案するようになった。

4. ユビキタス・データ分析環境の教育効果

4.1 対象年度の授業内容の違いについて

本稿で検討の対象とした調査実習は、新型コロナウイルスの流行に伴うロックダウンの翌年、2021年度に開講した講義で行われたものである。Pythonによるデータ分析環境の構築は、ロックダウンによる大学への入構制限及び遠隔講義の実施が開始された2020年度のことであったが、本報告での考察対象を2021年度の授業にしたのは次のような事情による。

まず、ロックダウン開始当初は遠隔講義の継続についての見通しがなく、本格的なデータ分析を行うことになる後期には対面講義への移行も考えられた。しかしながら、七月に入っても新型コロナウイルスの流行は続いており、大学側が対面講義の義務化を行わなかった。そのような状況の下で受講生に出校を強要すれば、通学過程および学内滞在中において受講生の罹患の可能性が増す危険がある。また、文系学生によるSPSS for Windowsの利用には、先に挙げたような障害があり、長時間にわたって罹患の危険を冒しながらそうした障害に立ち向かうことを強いられば、授業を放棄する学生が続出する可能性があった。Pythonによるユビキタス・データ分析環境の構築の当初の目的は、コロナに伴う遠隔授業という状況下で、対面授業時と遜色ない、あるいはそれ以上の教育効果を上げることであった。そして2021年度、ロックダウンがされた後においても、一定の条件のもとで遠隔授業の実施が容認された。これを機に、構築と運用の実績ができ、当初の想定以上の教育効果をあげた分析環境の使用を前提に授業計画を再構成し、授業の教育効果を評価できるようにした。以上が、本報告での考察対象を2021年度とした理由である。なおこの授業は、完全に遠隔で行なった。

さて、実習環境の効果を考察するには、二つの年度の教育効果を比較する必要があるが、ここではその十年前、SPSS for Windowsを用いて講義を行っていた2011年度の講義を対照させることにしたい。

ここで、比較対象を2019年度ではなく10年前のものにした理由は主に、(1)明星大学人文学部人間社会学科では、社会調査実習の講義が輪番制となっており、2019年度は小生の担当年度ではなかったことに加え、(2)2011年度から「自治体の住民を母集団に想定して標本抽出作業を行い、郵送調査で回収された調査票から作成したデータの分析を行う」方法を変更し、この年度は過去に収集したデータの二次分析にしたからである。その大きな理由は、受講生強い負担を軽減し、データ分析作業と分析結果の解釈にあてる時間を拡充することにあつた。また、(3)小生の担当年度の講義では、2011年度を最後に調査報告書の作成を取りやめ、それぞれの学生が作成したレポートをワードファイルの形式で提出させる形に変更した。これは、自らの関心と異なるテーマを課された学生の意欲が低下し、授業を放棄することを防ぐためであった。

つまり、二つの年度の授業内容は(1)学生各自の関心に基づく研究テーマの設定および二次データの分析によるレポート作成という点では共通しているが、(2)統計分析に用いるデータ分析環境という点で異なっていることから、両者の教育効果を比較することによって、分析環境の違いがデータ分析能力の育成に及ぼす影響が、明らかになるものと想定されるのである。

4.2 研究テーマの変化について

まず、学生個人の研究テーマについて見ていくことにしたい。2011年度の授業で分析対象としたデータは、2010年4月1日時点で満60歳以上の羽村市民を対象としたもので、調査テーマは「郊外住宅地のライフスタイルと社会意識」であった。学生たちは参考文献の学習を通してそれぞれ個別のテーマを設定したが、前年度の構成が社会意識・社会関係・社会参加・まちづくり・階層意識・地位達成アスピレーションという網羅的なものであったのに対して、二次分析を始めた2011年度の構成は、社会意識・社会関係・結婚や性的役割分業、教育意識、情報環境の活用、政治意識と、一冊の報告書にまとめるのが困難なほど多様な内容となっている。これは、1.調査票の作成と印刷、2.標本抽出と調査票の発送及び回収、3.膨大な量の回答のチェックとエディティングおよびコーディング、4.大量のデータの打ち込みとクリーニングなどの時間的・肉体的・精神的に過度な負担がなくなれば、文系の学生であっても、自由なテーマ設定および仮説の設定と検証作業の繰り返しを行うようになることを示唆している。

2021年の授業で分析対象としたデータは、2016年度に明星大学の在学学生を対象として行った情報行動についてのアンケート調査によって得られたものである。スマートフォンをはじめ、さまざまなモバイル機器を所持する学生が珍しくなくなり、それを前提とした多様なサービスが提供されるようになっていたから、情報行動や日常生活の情報化に関する意識についての質問項目は、多岐にわたった。また、東日本大震災の記憶が薄れる前であったので、災害時の情報行動に関する質問項目も少なくなかった。調査票には学生たちの日常生活および通学時間の過ごし方に関する質問項目と、学生向けに考案されたフェイスシートも掲載された。最終的に作成された調査票は、オーソドックスな質問質問に加えて、学生たちの関心に基づいた質問項目を多く取り入れたものとなった。

こうして得られたデータの二次分析は、1.既存データに関連させながら自らのテーマに基づく仮説を再構成し検証を重ねる必要があり、2.時代背景の異なる世代が置かれた背景についての情報を考慮しながら分析結果の解釈を行う必要があつた。必要な深掘りを加えながら、学生各自が納得できる分析結果を得るには、アイデアをすぐに試せる環境と、多くの分析時間が必要であったが、時間と場所を問わず簡単な操作で活用できる分析環境は、いつでもどこでもアジャイルなデータ分析を可能とした。その結果、学生たちが提出したレポートのテーマは、通学時間の情報行動、集団参加とSNS利用、日常生活での先端技術の活用意向という平時に関するものから、危機意識とSNS活用、災害時のSNSの信頼度という非常時に関するものまで、非常にバラエティに富んだものとなった。

4.3 データ分析スタイルの変化について

新変数の作成作業を教員が代行したことにより、自らの仮説を検証するための新変数作成の要望が増えた。この過程で変数合成についての理解が深まり、様々な新変数が提案され、分析に利用された。また、簡単にダブルクロス集計を行える環境を提供したことにより、第三変数を用いて自らの分析をより精緻化させて、レベルの高いレポートを作成する学生も現れるようになった。

5. おわりに

VUCA ワールドの到来を受け、リベラルアーツを活用しアジャイル型の仮説検証ができる「文系型データサイエンティスト」が求められるようになってきている。しかし、高校時点ですでに「数学への苦手意識」が浸透してしまった文系学生を、大学で文理融合人材に育てるのは難しい、とも言われる。小生は「文系学生に相応しい統計学教育のあり方」という視点から、「文系学生の世界観に基づく数理モデルの理解」のあり方を念頭に、統計学の教育コンテンツの再構築を行い、教育現場では大きな成果を挙げた³⁾ものの、教育内容のデザインが社会調査士課程の科目としては認められなくなり、この試みは潰えてしまった。

ただし、時代はSTEMではなくSTEAM、すなわち、客観的で数値化が可能なデータを数理モデルによって分析できる能力ではなく、リベラルアーツに基づいた発想および考察を合わせて意思決定できる能力。そして、量的調査の分析結果と質的調査の分析結果をより高いレベルで統合しながら考察できるセンスメイキング能力を持った人材を求めている。数学が苦手な文系学生であっても、教育の現場で様々な阻害要因を取り除けば、自由な発想で文系的なセンスを活かしながら研究テーマを設定し、既存の変数をもとに必要な新変数を作成して使用し、アジャイル的なデータ分析により検証するようになるのではないかと。そしてそれは、「数学嫌いの文系学生」を「文系的センスを持ったデータサイエンティスト」へと変える上で、非常に有力な方法論になるはずだ。これが、2021年度に小生が担当した授業の、大きなテーマだった。

標準化調査の実施に伴う様々な作業と、回収した調査票のチェックおよび電子データの作成などに伴う荷重な負担を取り除いてもなお、文系学生の前に立ちはだかる「統計パッケージを用いた仮説検証作業に必要な操作の習得」「新変数の合成のために必要となる操作の習得」「大学の施設を活用する上での時間的・空間的制約」という高い壁。小生が担当する授業では、ユビキタス・データ分析環境を提供してこれらの負担や制約を取り除けば、文系タイプの学生であっても、文系的センスを活かしたデータ分析に専念でき、より優れた能力を身につけることができるとの想定のもと、スマホでも活用できる分析環境を構築・提供したのであった。

こうした試みの効果は、先に見た通り、受講生の積極的な授業参加、新変数作成の提案、そして、提出されたレポートの質の向上という形となって現れた。また、これと同時に、途中で受講放棄をする学生の減少(データ分析期間における授業放棄は皆無)と、レポートの完成度の向上(誤字脱字および修辭法の両面において、大幅な修正が必要となるレポートは提出されなかった)が認められた。1.受講生各自が、先行研究に学びながら自由にテーマを設定できること。2.分析に必要な新変数について、自らの想定に基づき提案したものが、講義時間中に作成されすぐに検証できること。3.いつでもどこでも簡単な操作で仮説検証ができる環境を活用しながら納得できるまで分析できること。4.文系学生に相応しい数理モデルと検定理論の丁寧な解説に加え、5.こうした授業環境を提供したことごとが、学生たちを試行錯誤を伴うデータの分析に熱中させるとともに、精神的・時間的余裕を持ってレポート作成ができる状況をつくりだしたものと考えることができよう。

小生の授業は、社会調査士課程の中の、統計学の基礎や量的社会調査の実習に関わるものにはすぎず、こうした試みがどれだけ一般化できるかについては、様々な領域におけるケーススタディによって検証が必要と思われる。ただし、数学嫌いの文系学生から文系的センスを活かせるデータサイエンティストを育成する際に、

- 1.理系タイプの学生との世界観の違いを正しく理解した上で、文系学生に相応しい数学教育のあり方を(コンテンツ・教授法・評価方法を含めて)再構築すること
- 2.具体的なデータ分析の実習では、使用する分析手法について、文系学生でも理解できるように説明し、それに基づいたデータの分析ができるようにすること
- 3.実習授業では、受講生一人当たりの作業負担が過度にならないようにすること
- 4.統計パッケージ等の分析環境を、統計用語やプログラミングに疎い文系学生にとっても分かりやすく、使いやすいものにする(UIの問題だけではない)
- 5.時間的・空間的制約を受けず、いつでもどこでもアジャイルな分析ができるような、分析環境を構築し提供すること
- 6.調査データの分析を行う授業方法を、先行研究に依拠した仮説設定と統計パッケージの出力の解釈という形ではなく、先行研究は参考にするものの、基本的に学生の自由な発想に基づく仮説設定とアジャイルな仮説検証作業を行わせ、教員がそれをサポートする形に転換すること

が有効なことが示唆されたことは、重要であろう。文系学生の多くが高校時代までに数学に対する苦手意識を抱いていたとしても、大学での教育方法次第で、文系的センスを活かしたデータ分析ができる人材に育てることが、決して不可能ではないことを意味しているからである。変えるべきは、教育コンテンツと教育環境および教育法であり、重要なのは、それらの適切な再構築なのである。

注

- (1) この説明方法の根拠は、エヴェリット.B.S.(1980)にある。
- (2) ハッキング(1999)に着想を得て始めた研究の成果は、天野徹(2002)にまとめ、これをもとに文系学生のための統計学教育を再編成した教科書を作成した(天野徹(2006))。
- (3) 松尾由美・玉田和恵(2021)は、文系データサイエンティストを育てる上で、こうした方法が有効であるとしている。

参考文献

- (1) 天野徹著：“統計学の想像力—覚束ない未来のために—”，ハーベスト社 (2002)。
- (2) 天野徹著：“社会統計学へのアプローチ—思想と方法—”，ミネルヴァ書房 (2006)。
- (3) イアン・ハッキング著(石原英樹・重田園江訳)：“偶然を飼ひ慣らす”，木鐸社 (1999)。
- (4) エヴェリット.B.S.著(内山光哉監訳)：“質的データの解析—カイ二乗検定とその展開—”，新曜社 (1980)。
- (5) 符儒徳：“文系の学生向けデータサイエンス教育に関する一考察—モデルカリキュラムと認定プログラムに基づいて—”，開智国際大学紀要 第23号，pp.5-24 (2024)。
- (6) 松尾由美・玉田和恵：“私立・文系大学におけるデータサイエンス教育の課題”，江戸川大学紀要第31号，pp.249-255 (2021)。