

統計学におけるコンピュータ演習での データ駆動型教材の設計と実践について

信州大学医療技術短期大学部

鈴木治郎

〒390-8621 長野県松本市旭 3-1-1

TEL 0263-35-4599 Fax 0263-37-2370

e-mail: szkjiro@gipac.shinshu-u.ac.jp

概要

統計学入門課程に対して、コンピュータシミュレーションをもととして理解を図って行く大学初年級半年の学習課程を筆者は設計するとともに1997年4月から授業を実施し、それを教科書[1]に著した。コンピュータシミュレーションのためのツールは表計算ソフトの一つExcelであり、ワークシート上の操作のみで学習が完結するようにした。

ここで対象とした統計学は情報科の教材ではないが、(1)情報化社会において不可欠となった知識である、(2)情報科学同様十分に抽象的かつ論理的である、(3)実際にはコンピュータ利用が欠かせない、という情報科同様の条件を備えるものである。さらに予備知識は中学校数学程度を前提に設計しており、この教材の開発において出会った課題は、今後の高校情報科の具体的な展開を与える上での参考となる点が多いと考える。

教材設計の際に留意した点は、認知科学の知見を参考に、各課題が含んでいる統計学における論理的関係性を画面内に明示的に表現したこと、そしてパラメータの制御によりシミュレーションを変化させることおよびその繰り返しを通じて、ランダムな変化の中での不変部分として統計学を構成する論理的な性質を観察できるようにしたことである。

1. はじめに

コンピュータ利用による教材設計では、伝統的学習方法では困難であった事柄を一つ解消することができる。それは自由度の高いシミュレーションをうまく設計することができれば、伝統的学習方法では長い時間をかけて始めて理解が可能になる主要課題の特徴を多角的に提示することを可能とした点である。

しかしながら、たとえば数学教材の関数のグラフ化に際してコンピュータを活用した経験者の多くは、グラフィックスを楽しむだけでなく簡単なことなのに、その数学的意味合いを理解するためにはその表現した数学を理解していないとよくわからないという自己言及的状況が発生していることを言っている。

この例には、コンピュータを活用して数学理解を楽にしようという目的でありながら、コンピュータを活用した作成物の理解には数

学が分かっていないと難しい、という「卵が先か、ニワトリが先か」的自己矛盾が現れており、こうした問題は数学教育に限らずコンピュータ活用の場面ではよく現れるものである。

2. 情報科との共通性

高校情報科と統計学には

1. 情報化社会に生活する上で不可欠となった知識である
2. 情報科学同様十分に抽象的かつ論理的である
3. 実際にはコンピュータ利用が欠かせない

という共通性がみられる。いかにその根拠を述べて行く。

2.1 日常生活における必要性

高校情報科新設の必要性に関する背景をひ

とことと言えば、その科目が情報化社会で生きる私たちにとって不可欠な内容を有するからである。

本説で対象としている統計学についても、日常生活では

- 受動喫煙が肺ガンにつながる
- 内閣支持率が 47.3%であった

のように、推測統計に関する知識抜きにはその意味を正確には把握しがたい情報に接することが珍しくなくなっている。こうした情報が正しく把握されていないことの証拠には、たとえば報道機関によって「世論調査によると内閣支持率が 0.3%上昇した」のような報道が珍しくないことを通じて伺える。この例では世論調査における誤差範囲内(約 2000 人の調査では誤差は約±1%)の測定値の変化を「上昇した」と言っている。この例を含めて平均値だけが報道され、それを比較するのに必要な標準誤差が報道されない状況に違和感を覚えるような教育が考えられてしかるべきである。

2.2 統計学の抽象性と論理性

統計学は数学(確率論)を基礎に記述されており、情報科学を構成する多くの要素が抽象数学で記述されている状況と似ている。つまり統計学の基礎をなす事柄である「確率」等が数学的に定義されているのである。「確率」等の概念を例示することと、その抽象的概念を理解することの間にあるギャップをいかに埋めるかが教育における課題となる。

2.3 コンピュータ利用の不可欠さ

通常の統計学入門レベルの知識を利用する上で必要となる計算力は中学校 3 年レベルの数学知識(平方根を含む数式の計算)である。しかしながらデータの統計処理に現れる計算は総和記号 Σ を用いて表現するほうが自然なように単純な計算の繰り返しを多く必要とする。また実際のデータを扱う際には、計算過程で小数点以下に多くの桁の現れる数値

を扱うことになる。これら特徴はコンピュータ利用が自然であることを示している。さらに医学や心理学等の高度な統計処理を必要とする領域における統計処理にはコンピュータを使うことが当然のこととなっている。

3. 教材設計の基本方針

教材設計に当たっては

1. 日常生活に現れる題材と統計学の主張との対応を自然に理解できるようにする
 2. 数学的主張が自然に観察できることを示して理解を図る
- を基本に据えた。これら視点について具体的に説明する。

3.1 日常生活に現れる題材との対応

日常生活に現れる題材の中で、統計学の啓蒙書でもよく扱われているものから

- 世論調査等の比率の推定
- 不確かな事柄への対処等、確率的状況の評価に使われる確率の意味合いの理解
- 大数の法則が日常の中の確率的現象にどのような形で現れるか

を理解目標に設定した。理解を図って行く目標をまず設定し、統計学入門の学習コースをこれら目標に向かって再配置するというトップダウン的設計とした。

数学的理解を踏んで行くと十分な時間がないから統計数学の学習で終わる、あるいは統計計算処理方法を天下り的に認めて数学的な意味もよくわからないままに演習を行なう、という従来の教科書に見られた立場ではなく、目標とする課題が数学的に自然な主張であることが理解できるように授業展開の限られた時間数の中にコンピュータシミュレーションを並べて行くようにした。この方法はトップダウン的と称してよいだろう。

必要な時間内に課題を配置するために、統計学の標準的コースの記述では不可欠なのに切り捨てたもの(例:チェビシェフの不等式)

もあるし、逆にそうしたコースでは 1 行ぐらいしか触れられていない（例：確率変数の実現値とはどういうことか）のに、いくつものシミュレーションに展開したものもある。これらアプローチは齊藤[6]のいう工学的アプローチでもある。つまり目標の解決を第一に、現在利用可能な資源（既存の統計学コースにおける蓄積、啓蒙書およびコンピュータ計算）をいかに配置するかによって、課題の解決を図ること、また、それに当たっては方法間の矛盾も生じるが、対象が複雑ゆえできあがった設計がうまく働くかによって評価される、というものである。

統計学入門コースを述べるのに、ここではコンピュータシミュレーションによる体験を通じて自然に観察される現象として統計学の課題を扱い、観察を通じて発見できた性質を命題として確立して行った。

コンピュータ利用ではトピック中心の例示で面白くというアプローチも考えられるが、ここでは基礎的な学問の学習にのっとった勉強を通じて学力を高めていくという伝統的に行われてきた初等中等教育の視点を大事にて構成している。それでありながら、生徒の卒業後の社会生活に対する責任感をトップダウン的な設計として表現した。

3.2 数学的主張の観察

数学の学習においては

- 数学的概念を理解しないと先に進めない
- 数学的概念を操作するための計算に習熟しないと先に進めない

という学習上の困難があったが、それをコンピュータを活用した計算によって適宜スキップすることは可能になった。統計学においても適当なコンピュータアプリケーション利用のもとで容易な学習を売り物にした教科書が現れるようになった。そうした教科書のほとんどはコンピュータに計算代行させるためのマニュアル本に近い体裁となっている[9]。

一方で、数学の運用能力の高い人たちにあ

っては、数学の証明をよく理解しているという事実が共通してみられる。そのため、統計数学を学習する際に出会う証明自体の理解をコンピュータ活用の主対象とすることを試みた。

そして統計数学に現れる主張の多くは数えることおよび確率収束に関するものである。そこで、この試みの実現のために確率収束に関する事柄を確率変数の実現値の振る舞いから相対頻度の近似がどのように変化するかを主体としたシミュレーションに構築した。

4. 認知科学からの知見

複雑な現象の理解と学習に関する事柄はほとんど未知な現状であるが、人がどのような理解の仕方をするか、あるいはどのようなデザインが限りやすいかということについては、認知科学により多くのことが知られてきている。ここでは、そうした知見と、それらを教材設計にどのように反映させたかを述べて行く。初めの 5 項の区分は[7]にしたがった。

4.1 論理的理解

人は実験的課題の多くで論理的に正しい結論を導けることが知られている一方で、論理的課題としては同等であっても、題材とした対象に関して、どのような経験が積まれているかによって正答率の変化する場合のあることが知られている。

数学の学習法の大半は、上記の正しい論理性を人がもつことを前提に組み立てられているが、その一方で学習のつまずきの大半は誤った経験の定着によるものである。

ところで、統計学の基礎となるランダムさそのものについては数学的に「何ら法則性が見られないもの」としか言いようがなく、平均操作を通じて初めて等確率性を目にすることができるとするのが、統計学の標準的コースの現状である。

そこで、ランダムさに関する経験をコンピュータシミュレーションを通じて十分に経験を積めるように教材を設計した。

とくにランダムさに見られる等確率性と小数の現象では見られない一様性とを混同している状況の改善がみられたと感じている。

4.2 ルールの集合による推論

「もし～ならば～をせよ」というルールをたくさん使い分けて人はさまざまな判断を行なうという立場である。

高度な統計学の手法を必要とする医学研究者と筆者は日常的に交流を行なう職場にいる。そうした交流において、彼らはこの「もし～ならば～をせよ」の知識として研究活動における統計学を把握しているものの、自分の分析したデータに対して「もし～ならば」部分をどのように考えたらいいかわからないことに最大の問題点を抱えてた。

そこで、統計学に関するコンピュータシミュレーションを展開するに際しては、統計学上の命題の仮定部分がどのように反映されるかをできる限りわかりやすく提示できるよう設計した。

4.3 概念による理解

人は物事の理解において心の中に形成された概念モデルに合わせて、観察した現象を理解すると考える立場である。

そこで、ランダムさに関する概念をコンピュータシミュレーションで得られる経験を通じて形成して行けるよう、ランダムさのシミュレーションの利用に一貫性をもたせた設計を行なった。

また、統計解析はランダムさに関する実験結果と、それに対応する統計学の命題を対比させることだという概念が得られるように、統計学上の命題の扱いに一貫性をもたせた。

4.4 類推による理解

人はすでにもっているモデルとの類推により物事の理解をはかっていくという立場である。

教材設計に当たっては、乱数実験の1回1回が実際のデータ収集や観察の1回に相当

するという類推を通じて、統計的推測の方法論が得られるという理解を促すよう記述した。

4.5 イメージによる理解

いろいろな事象に対して視覚的イメージをもとに理解していくという立場である。

教材設計にあたっては乱数実験結果のグラフを何回も何回も実験の繰り返しの度に観察することを通じて、ランダムさに関する視覚的イメージを形成してくれることを期待して課題の設計を行なった。

学習評価に当たって、ここであげた5つの立場の内、最もことばで表現しづらい部分であり、授業中の学生との会話を通じて期待した事柄を獲得できた学生が多くいると感じている。

4.6 確率・統計的現象に関する誤認

トヴェルスキーとカーネマンにより確率的現象に関して私たちがもっている誤認の原因が考察されている[8]。教材設計の際に参考にしたものを以下にあげる。

- 少数の法則：大数の法則が少数例にも現れるはずだとするもの
- 代表性仮説：標本には母集団の重要な特性がすべて表れているはずだとするもの
- 利用しやすさ仮説：ある事象に対して確率を推定するときに想起しやすい事象を参考にするというもの
- 投錨と調整仮説：初めに立てた確率の推定に引きずられて、それを修正すべき情報が得られても十分な調整ができないというもの

4.7 デザインの原理

ノーマン[4]はこれら人の理解の仕方、あるいは誤認のしやすさに関する認知科学の知見をもとに、とくにコンピュータ利用に関するデザインの原理を次のように述べている。

- いついかなるときにも、その時点でどんな行為をすることができるのかを簡

単にわかるようにしておくこと（制約を利用する）。

- 対象を目に見えるようにすること。システムの概念モデルや、他にはどんな行為を行なうことができるか、そして、行為の結果なども目に見えるようにすること。
- システムの現在の状態を評価しやすくしておくこと
- 意図とその実現に必要な行為の対応関係、行為とその結果起こることとの対応関係、目に見える情報とシステムの状態の解釈の対応関係などにおいて、自然な対応づけを尊重し、それに従うこと。

これらの点をできる限り実現できるよう教材設計に留意した。教育行為の場合、こうした設計の最も難しい点は、教育の過程の進行に応じて利用者＝学習者がもつ概念モデルを次々に変化させながら、よりよい概念モデルに近づけて行くこと自体が教育行為だという点である。

制約を利用するということは教え過ぎない、なるべく基本的な知識を使い回しせざるを得ない状況に生徒を置くことだと考えて教材設計を行なった。

たとえば、グラフの作成および観察においては、度数分布表の階級を常に横軸にとって、縦軸に度数をとるという一貫性を与えることによって、理論分布における確率変数の実現値との対応があるものとして、グラフ化された現象を常に評価するようにもって行った。

この見方を維持するために、統計学入門でよく扱う事項の一つ「相関」は除外した。

ところで、こうしたものの見方の固定化が自由な発想の妨げになるという考えもあるだろう。その点に関しては、数学能力のすでにある学生に対しては、これら観察からいかにして通常の統計学の教科書で学習していくべきかを随時促すよう、記述に注意した。また、基礎学力の段階で自由な発想を強調することは無知を放置することとあまり変わらないこと

も承知しておく必要がある。

このような配慮のもと、おそらくこの教科書以上には統計学を学習することなく社会生活を送るだろう人たちに、推測統計で得られる知識を日常的にどう判断するかを経験を与えるものとして、教科書を構成した。

とくに高校卒業後は数学や情報科学を学習することがないだろう学生を学習対象者とした場合には、考えればわかるものだという漠然とした信念は教育側および学習者側にとって教科に対する嫌悪感を生み出すリスクのほうが大きいのではないだろうか。考えればわかる、という多大な時間を要する学習法は趣味講座としての生涯学習の中にしか得られないのである。

5. 視覚的デザインの具体化

以上述べてきた特徴の具体的な実現に当たっては

- 操作のステップ一つ一つをなるべく学習課題に対応させる
- 操作のステップごとに画面表示の変化として確認でき、それらが学習課題の何に対応するかも確認できる

ことを常に実現できるよう表計算上での統計学シミュレーションを編成した。これら特徴は、演習において学習者が目にするのできたデータをもとに統計学の数学的側面の数々を確認し、逆にそれを数学的諸性質として演習にフィードバックを図って行くものである。

この方法を採用した結果、

- 画面上のデータにもとづいて、次に行なうべき操作が何であるかを知る手がかりを多く与えること
- 画面上のデータにもとづく表現を数学的にどう解釈して行くか考えさせること

がほとんどの課題で実現されており、それをここではデータ駆動型の教材と呼ぶ。データに対する画面上の表示から学習者が受ける印象を、伝統的な学習で得られる数学的諸性質

に対していかに結び付けるかが教材設計のポイントとなった。

これらの点を二項分布に関する以下の例について述べる。図1はベルヌーイ試行を理解するためのシミュレーションであり、1回の試行における成功率 p の値の変化が成功1と失敗0の並び方の様子にいかにかに反映されるかの観察を行なう。なお、図上の矢印は Excel のワークシート分析ツールにより表示させた。

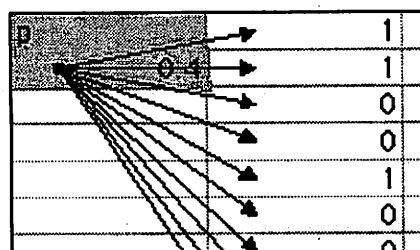


図1. パラメータと実験結果の対応

図2はこのベルヌーイ試行6回分の0と1の並びを合計することで、6回の試行中の成功数を与えている。この成功数を図3で度数分布表に集計させることにより、二項分布の観察に結び付けている。

p	0.4						合計
1	0	0	0	1	1	3	
0	0	0	1	1	0	2	
1	0	1	1	1	0	4	
0	0	0	0	1	1	2	
1	0	1	0	1	0	3	

図2. 実験結果の数え方と解釈

図3では成功数の集計度数に対して相対度数を計算し、相対頻度を二項分布における確率と結び付けるための計算を行なっている。さらに、図からは明らかではないが、度数分布表の集計に用いたものと同じ成功数、ベルヌーイ試行のパラメータに用いたものと同じ値を成功率に用いて Excel の二項分布関数の値を計算させている。

成功数	度数	相対度数	二項分布
0	1	0.033	0.0467
1	8	0.267	0.1866
2	5	0.167	0.311
3	12	0.400	0.2765
4	4	0.133	0.1382
5	0	0.000	0.0369
6	0	0.000	0.0041
合計	30		

図3. 実験結果の度数分布表

図4ではこれら相対度数と二項分布を一つのグラフに表示させてある。

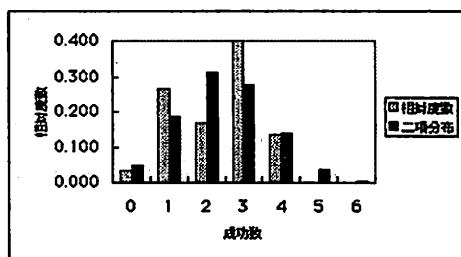


図4. 度数分布表のグラフ化

乱数実験の繰り返しの度に相対度数のグラフは高さが変化するが、二項分布を表すグラフは不変である。このような視覚的観察を通じて乱数実験を表すグラフの変化の背後に理論分布の存在を感じさせるように図った。

なお、実際の教材の展開においては、理論上の二項分布を計算させない段階でグラフを作成し、学生たちにその観察を通じた予想を行なわせている。

さらに各課題を展開するに当たっては、コンピュータの具体的操作を徐々に減らすことにより、操作に関する知識はいわばパッケージ化された知識として用いて行くように促し、最後には数学的記述のみで対応するシミュレーションを理解できるように図った。

6. 授業評価

以上のような発見型の進行を伴った学習課程において、適切な評価手段を見出すことは容易ではない。ここで考えたレポート問題は、

たとえば「二項分布を与えるベルヌーイ試行の列を与えると考えられる身の回りの現象をあげるとともに、その理由を説明せよ」のようなものを用いた。

こうした課題の展開に表計算ソフトの Excel 自体は不可欠ではないが、ほとんどの課題をマウス操作のみで行なえること、またベストセラーの表計算ソフトであるためにそれを使いこなしたいという学習のための動機をもつ者が多く見られたという特徴があった。

7. 何を取り入れなかったか

以下にあげるような事項は取り入れないように配慮した。

7.1 グラフの視覚効果改善のための努力

コンピュータ利用の授業では学生が最も楽しむことのひとつである。時間的にきりがいいこと、個人的に楽しんだほうがよいこと、という判断のもと教科書の記述としては、視覚的要素の方法を Excel の扱い方の一項として置くにとどめた。実際の授業では学生個別の状況に応じて活用の示唆は行なっている。

7.2 複雑な手順の課題

一連の複雑な操作を通じて、手順書通りに進めれば学生一人ではけっしてできないようなこともできる、という課題は置かなかった。

一見複雑そうな課題を扱う場合には、それは関係例題の発展として考えれば容易に実行できるように置いた。

7.3 学生の学習活動

学生が教科書を活用しながら学習を進めて行く中で、気になった事柄をあげておく。いずれの状況も 2~3 割の学生に見られたものである。

- 電卓として表計算ソフトを十分に使えない。転記の必要のない計算用紙として便利さを十分に活用している学生のいる一方で、たとえば「平面上の 2 点の座標が与えられたときに、その 2 点

間の距離の計算」を求めるような場合に、一つのセルに長い式をいきなり作り上げようとする学生も少なくなかった。この例でいえば、平方根関数が SQRT である、という予備知識および数学表現での数式は示した上で生じている問題である。

- 画面に表示された実行結果が教科書の記述と違った場合には実行結果を無視して先に進んでしまう
- 教科書の索引を引かない
- 操作指示的文章の目的語を厳密なものとして読まない
- うまく実行できなかった課題をやり直してみない

コンピュータ利用に関しては、入学までに全く触れたことのない学生は少なくなっているものの、エラー等に悩まされる状況で使い続けるまでのプレッシャーの中でコンピュータを利用するのは初めてという学生が大半である。そうした状況下で授業を行なってみて

- ソフトウェア操作の難易度：操作技術を限定したので慣れるにしたがって問題はなくなった
- ソフトウェアのくせ：時間をかけるにしたがって慣れてしまう

ものであり、初心者向けパソコン雑誌で目に付く初心者のもつ恐怖感に関する問題は実際には問題とならなかった。

授業が進行しても基本的な学習態度として問題なのは、[分数のできない大学生]でも述べられている国語能力上の問題である。エラーレポートが正確にできない。人間関係と違って、コンピュータはしばらく我慢していればエラーは消えてなくなるということはない。

8. おわりに

ここで述べた教材設計において、統計学という教材に強く依存している特徴は乱数実験の繰り返しだけでも十分に変化に富んだシミュレーションを与えることができた点である。それ以外の特徴は表計算ソフトの特長であ

るコピー機能を利用してたくさんの計算を容易に実現させることや、グラフ化の機能については数理的要素のある科目であれば十分に生かすことができる。たとえば環境問題の学習に CO₂ の排出量の計算や化学式の原子量の計算など、パラメータの制御によって変化を確認して行く問題は表計算ソフトの特性に馴染むだろう。

また、学習者の操作と画面を通してのプログラム側の反応を、なるべくブラックボックスではなく見せる工夫することにより、抽象的な方法に関する信頼感を深めることのできる可能性を示唆できたものと信ずる。

参考文献

- 1) 鈴木治郎：「Excel で実験する統計学」，p.339，アジソン・ウェスレイ・ジャパン(1998)
- 2) 鈴木治郎：「コンピュータシミュレーションで育てる統計的直観」，コンピュータ&エデュケーション，Vol. 4，pp. 39-43 (1998)
- 3) 鈴木治郎：「GIF アニメーションによる統計学博物館の試み」，1999 年情報学シンポジウム講演論文集，pp. 73-79 (1999)
- 4) D. A. ノーマン：「誰のためのデザイン？」，p. 403，新曜社(1990, 原著 1988)
- 5) 石川幹人：「サイコロと Excel で体感する統計解析」，p. 137，共立出版(1997)
- 6) 齊藤了文：「〈ものづくり〉と複雑系」，p.283，講談社(1998)
- 7) Paul Thagard：「マインド 認知科学入門 (松原仁監訳)」，p.238，共立出版(1999, 原著 1996)
- 8) 佐伯胖：「認知科学の方法」，p.258，岩波書店(1986)
- 9) 内田治：「すぐわかる Excel による統計解析」，p.208，東京図書(1996)