

高校生が楽しく学べる仮説検定と相関・因果

シミュレーションから始まる”情報科”統計教育

千葉県 日出学園中学校・高等学校 武善 紀之

新学習指導要領ではデータサイエンスが大きな目玉となっている。現行課程の内容でも統計に対しては苦手意識を持っている生徒は多い。今回は楽しく統計の本質を学べるような実践として、数式を一切使わない仮説検定の概念習得及び実習、相関関係・因果関係の発見グループワークを行った。結果、生徒の興味・関心を維持しながら、統計に対する理解を大幅に深めることができた。その実践の流れと成果について報告する。

1. 実践の背景ときっかけ

統計分野は苦手意識の強い生徒が殊更多く、計算が面倒、性格が悪くなりそうといった負のイメージをよく持たれてしまう。これを払拭するため現課程ではクラスアンケートの調査結果を1枚のポスターにまとめる実習を行っていた。生徒も楽しんで取り組んでいたが、単なる何でもアンケートに陥ってしまうことが課題であった。

新学習指導要領ではデータサイエンスが1つの目玉となっており、情報Iと並んで、数学Iでも仮説検定の考え方や相関・因果まで扱うこととなった。これに対し数学=理屈、情報=ソフト活用と棲み分けてしまうと、情報=PC教室の二の舞に陥るのではと、危惧を抱いている。

情報科としては理屈も徐々に理解させながら、何より楽しく統計を学び、体験して欲しい。今回は今までの課題と共に情報Iを意識してクラスアンケート実習を改良し、数式を一切使わず「仮説検定」と「相関・因果」へ繋げた実践を報告する。

2. 実践1:統計的仮説検定実習(6コマ)

2.1 導入:シミュレーションに基づく推論

仮説検定においては確率密度関数や数式指導が障壁となる。今回は Doi(2018)の開発した教材^①を用いて、理論の説明を行った。この教材は「赤ちゃんに善悪はあるか?」という心理実験で生徒の興味を惹き、数式無しで仮説検定を理解させる。

“ある映像を視聴した赤ちゃんに図形の選択を行わせると、16回中14回図形Aを選んだ。この時、赤ちゃんはデタラメに図形を選んでいるか(偶然)、確固たる意志(必然)で図形を選んでいるか?”

テーマの提示後、生徒にコインを配る。「デタラメ」という仮説の下、赤ちゃんをシミュレートという設定で、16回のコイントスを全員が行い、クラス全体の確率分布を作り上げる。その上で「偶然ではあり得ないことが起きている」→「仮説が間違い」という仮説検定の考え方を理解させる。

更にその後、コイントスをもっと大きな試行回

数で Web^②上にて試す。シミュレーションを2度繰り返すことにより数式を一切使わず、仮説検定の考え方を無理なく掴むことが出来る。

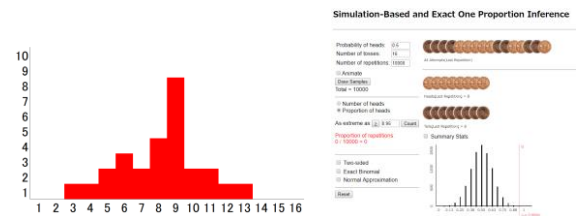


図1 2種類のシミュレーション

2.2 実践: Fisher の正確検定

Doi の実践は二項検定であり理解しやすいが、実習テーマとしてはやや面白みに欠けてしまう。

そこで同様に正確確率検定である Fisher の正確検定を実習では用いた。この検定は 2×2 のクロス集計表に対して期待値計算でその表の珍しさを算出し、そこから p 値を求める。そのイメージは SBI の導入だけで十分可能であり、検定も R を用いれば、たった 2 行の命令で行うことが出来る。

運動部に入ると、
運動が好きになる？

```
> mx=matrix(c(12, 6, 5, 14),  
nrow=2, byrow=T)  
> fisher.test(mx)
```

Fisher's Exact
p-value = 0.02171

↓

p<.05より有意

	運動好き	運動嫌い
運動部	12	6
それ以外	5	14

図2 直接確率計算のイメージとRコマンド

この手法では、 2×2 のクロス集計表に落とし込めない設問は実施できない。中には再提出を促す必要のある生徒もいたが、むしろ今までのアンケート調査実習よりも興味深いテーマが多く出たように思う。(以下テーマ例を記載)

- ・好物を食べる順序と兄弟の存在の関係
- ・SNS の好み(Twitter/Instagram)と性別の関係
- ・小雨の際の傘差し頻度と性別の関係

2.3 配当時間、授業の流れ、結果

SBI 講義 (1 コマ) 後、生徒は仮説を設定し、アンケート項目を作成した。上記の例では、「①運

動部所属で運動好き」「②運動部所属で運動嫌い」「③運動部以外で運動好き」「④運動部以外で運動嫌い」の4択アンケートを行えばよく、Google Form を活用すれば、データ収集は比較的容易である。可能な限り「よく食べる」などの項目は「週に3回以上食べる」と定量的に表現するよう心掛けたが、考察の執筆までを踏まえ、あえて主観のレベルに留めた調査もあった。

アンケートへの回答は宿題として、次の時間に集計、R で検定を実施した(0.5 コマ)。別クラスにも回答協力を仰ぎ、54名の回答データを用いたところ、全28名の履修者のうち、有意差が出たのは11名であった(いわゆる有意傾向である $p < .10$ を認めた)。その後、プレゼン作成(2.5 コマ)、プレゼン(1人3分間,2コマ)の計6コマで実施した。

3. 実践2: 相関? 因果? GW(3コマ)

3.1 因果関係を決める3つの規準

では有意差が出れば、そこには本当に因果関係があるのか。授業では次の例をいつも紹介する。

- ・小学校で無作為に30人の児童を抽出し、算数のテストを実施すると、身長が高いほど算数の点数が高くなる。つまり、身長が高いことによって算数の成績が向上する。【交絡因子「学年」】

因果関係の証明には共変関係に加えて、時間的順序関係ともっともらしい他の原因の排除という3つの規準が必要であるとされている⁽³⁾。

今までは例を紹介するだけであったが、今回は自分たちが導き出した結論に、本当に因果関係があったのかを考えさせるGWを実践した。

3.2 相関・因果関係図式化の方法

題材は慶應義塾大学(以下慶應)の入試問題から着想を得た。教科「情報」入試以前から、慶應は商学部、総合政策学部で相関と因果に関する問題を毎年のように出題している。例えば過去には、「コーヒーを飲むと生産性が上がる」という疑似相関に対し考えられる要因を列挙し、構造図を書き出して、分析を深めるという問題が扱われた。

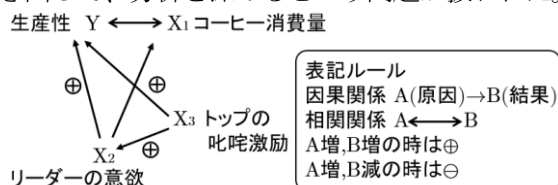


図3 入試問題図⁽⁴⁾を元に作成した図

この方法で自分たちの研究についても図式化し、考察の妥当性について発表するという実習を行った(生徒の研究は質的変数同士の比較なので「相関」ではなく正確には「連関」)。

3.3 配当時間、授業の流れ、結果

4人グループを7班作成し、慶應の問題を実際に解き(0.5 コマ)、その上で統計実習にて有意差が

出た5名の生徒の結果と結論から1つを選択して要因の列挙&図式化(1.5 コマ)、班ごとに発表を行い、もっとも説得力のある説明&図式に対して投票も行った(1 コマ)。同一テーマに対しても様々な解釈が生まれ、各自の経験を元に、図式化も発表も大いに盛り上がった。

テーマ:「兄弟の存在」が「子ども好き」を生む?

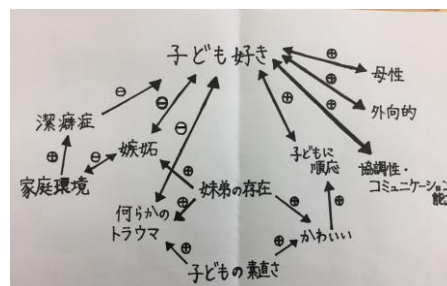


図4 生徒作成図の一例

4. 振り返りと今後の展望

4.1 生徒の定着度と授業への満足度

学期末試験において、仮説の設定や相関因果の違いを説明する問いでは正答率60%以上となった。もっとも正答率が低かったものは検定の非対称性を問う問題で、27%となり課題が残った。授業後に5段階で自己評価を行ったところ、平均は難易度(1易-5難)3.3、理解度(1低-5高)4.1、興味関心(1低-5高)4.3、日常におけるデータへの視点変化(0無-5有)4.1となった。

4.2 おわりに

授業後、生徒から「今回の実習で情報リテラシーは情報をむやみに疑うことではないと実感した」というコメントがあった。情報モラルは闇雲に疑うことではなく、情報の科学的理解に基づいた疑うコツの指導であることが以前から強く叫ばれている。メディアリテラシー領域における疑うコツがおそらく統計教育であり、目前の数式に惑わされたまま、本質を見失ってしまってもったいない。今回の実践を発展させ、情報Iにおける統計教育の方向性を引き続き模索していきたい。

引用文献・引用サイト

- (1) Jimmy A.Do, Active Learning Lectures for Statistical Understanding, 第6回情報教育研究会 in 江戸川大学 (2018)
- (2) <http://www.rossmanchance.com/applets/OneProp/OneProp.htm>
- (3) E.B.ゼックミスタほか, クリティカルシンキング 入門篇, 北大氏書房 (1996)
- (4) 慶應義塾大学 総合政策学部 2017年度小論文参考文献
- (5) 大貫和則, 統計リテラシーを育成するアンケート調査実習の実践と課題, 第4回全国高等学校情報教育研究大会 (2011)