

第4学年3組 データサイエンス I (DS I) 学習指導計画

指導者 中田 雅之

日時 2021年10月21日(木) ～ 2022年2月10日(木) 2限

場所 情報教室1

1 単元の設定にあたって

(1) 生徒について(「意識調査をもとに」)

単元設定に先立ち、生徒に対して行ったアンケート(表1)の概要について述べる。

表1: アンケート結果

(2021年10月11日実施, 第4学年生徒対象, 108名回答)

「そう思う」を「1」, 「そう思わない」を「5」とする5件法によるアンケート

		1	2	3	4	5	平均
1	物事の全体像に意識を向けることができる	9	41	46	12	0	2.56
2	統計的な図やグラフが目に残る	18	38	36	15	1	2.47
3	データの分布の様子を把握するために統計的な図やグラフを作成することは、意味のあることである	60	39	7	1	1	1.56
4	統計量(平均や四分位数, 標準偏差など)の意味が説明できる	26	57	18	7	0	2.06
5	統計量(平均や四分位数, 標準偏差など)を計算することができる	15	60	25	8	0	2.24
6	データから母集団の分布の様子に意識を向けることができる	10	42	37	17	2	2.62
7	データの分布の様子を数量化して考えることができる	4	42	45	17	0	2.69
8	データの分布の様子を数量化することには意味がある	37	48	14	7	2	1.97
9	統計は個々のデータの値について何も説明しない曖昧な道具である	2	4	27	54	21	3.81
10	統計は数学的に理路整然としない曖昧な道具である	0	1	18	58	31	4.10
11	統計は難しい数学を使うので利用したくない	1	13	39	42	13	3.49
12	統計は、数学的に裏付けされた理論に基づき、データの全体の様子を分析する有用な道具である	28	62	13	5	0	1.95
13	さらに進んだ統計の手法を学びたい	23	39	34	10	2	2.34
14	授業で学んだ分析手法を使って様々な事例の分析がしたい	25	47	20	12	4	2.29
15	統計の分析手法さえ理解できれば、背後の数学的理解は必要ない	4	5	15	50	34	3.97
16	データの分析にコンピュータを利用することは楽しい	30	36	30	10	2	2.24
17	データの分析にコンピュータを利用することには抵抗がある	5	14	16	42	31	3.74
18	Excel等の表計算ソフトを利用して統計量の計算ができる	11	47	28	17	5	2.61
19	Pythonを利用して統計量の計算ができる	2	15	25	48	18	3.60
20	Excel等の表計算ソフトには抵抗がないが、pythonを利用することには抵抗がある	12	36	21	25	14	2.94
21	コンピュータを利用することは、数学の理解にもつながる	29	43	29	7	0	2.13
22	コンピュータは素手では困難な計算を素早く行うことができる便利な道具である	66	30	10	2	0	1.52

まず、「設問3」「設問8」「設問13」「設問14」等の結果から、生徒は統計を学ぶことの意義について理解しているものと思われる。特に、「設問13」「設問14」の結果から、統計の理論を学ぶだけではなく、実際に活用することに意欲のある生徒が多いことには、注目しておくべきである。

統計と数学と計算機の関係については、「設問12」の結果から、生徒の意識下で、統計の背後に数学が用いられていることについて、同意がなされているとみられる。また、「設問21」の結果からは、計算機の利用と数学の理解についての関係性を見出しているとみられる。これらの結果から、生徒は統計、数学、計算機の間における結びつきについて、生徒なりの目線でこれらの関係性を認識していると思われる。他方、「設問19」や「設問20」の回答の様子から、統計の分析において計算機を用いることにつ

いての不安を抱えている生徒が一定数いることが窺える。これまでも使用した経験のある Excel と比べ、プログラミング要素をもつ Python への抵抗を感じている生徒が少なくないことについては、配慮が必要であろう。

授業者にとって意外な結果であったのが、「設問 9」と「設問 10」の結果である。「設問 9」について、授業者はこの点が統計の本質であり、ややもすると学習者にとって統計に近寄りやすい印象を与える点であると考えている。また、「設問 10」については、授業者自身が統計に対してかつて抱いていた感覚でもある。しかし、この結果からは、生徒は想像しているほどには統計に対して曖昧さを感じていないことが分かった。生徒の方が、統計による推論の仕組みについて、より深く理解しているのか、あるいは、生徒は統計に対して盲信をする傾向にあるのか、この調査だけでは分からない。

(2) 学習方法について

本校では昨年度より Google の提供する学習支援システム Google Workspace (Google1, 以下 GW) を利用しており、生徒は GW を利用して授業に関する連絡を受け取ったり、課題の提出を行ったりしている。GW の利点は、生徒の利用する端末に依存せず共通のプラットフォームを利用することができることにある。共通の環境で学習を進めることができるため、生徒同士で疑問点を気軽に訊き合うこともできる。

DS の授業においても、生徒はこれまでに GW に付属する表計算ソフトを利用してデータの整理や平均、分散などの統計量の計算を経験してきているため、授業内で GW を用いた作業を行うこと自体には慣れている。表計算ソフトや関数電卓を用いた統計の授業実践については、すでに多く試みられている（例えば、半田、清水、2021）。

今回、授業における統計分析のためのツールとして、スクリプト言語 Python を利用する。GW では Python による開発環境 Google Colaboratory (Google2, 以下 Colab) が利用できる。Python を利用することについて、以下の狙いがある。

- イ) 同時期に情報科では Python を利用したプログラミングの授業が進められており、学習の相乗効果が見込まれる。両授業で Colab を利用するため、学習が容易になる。
- ロ) 来年度データサイエンスⅡ (DSⅡ) では機械学習に関する内容を取り扱う見込みである。Python は機械学習のためのライブラリが充実している。この時期に Python の基本的な扱いに慣れておくことで、来年度の DSⅡ の学習を円滑に進められることが期待される。
- ハ) Python は操作をスクリプトとして記述するため、どのような分析を行ったのか、後から確認することができる。また、これは指導のしやすさにも繋がる。

Excel を筆頭とする表計算ソフトはデータを見ながら分析を行うことができ、操作も直観的である。直観的な扱いやすさの点では、Python は明らかに表計算ソフトに劣るだろう。しかし、上記の点を鑑み、授業では Python を採用する。スクリプトの記述に不慣れな生徒が大半であるため、授業で用いるワークシートは補助的なコメントを多く付けるなどの配慮が必要である。徐々に補助輪を外すように、ワークシートの学習フローをデザインしたい。

統計を学ぶにあたり、生徒が自分で興味のある題材を選択し、データをとって分析することが重要である。そのため、授業では活動と発表の時間を設定し、4 人規模の小集団を編成し、お題に沿った分析の活動に取り組むことを通して、学んだ統計手法の理解の深化に結び付けることができればよいと考える。この活動を繰り返すことにより、例えば日常生活や社会の事象から数理現象を見出す意識付け、数学化のサイクルを回す原動力などに貢献することも期待できよう。

(3) 教材について

本授業では、推測統計を扱う。本校では、総合的な学習・探究の時間において、統計を利用して分析

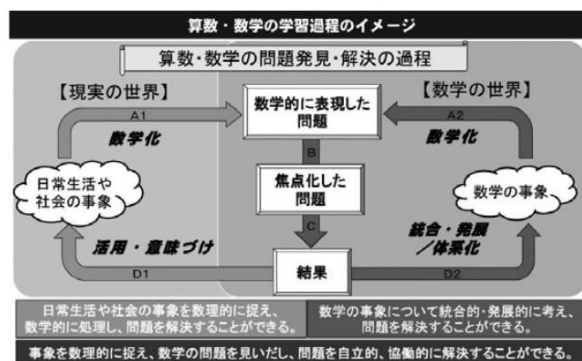


図 1 (文部科学省, 2019)

や考察を行う場面が想定される。そのため、授業においては、数学の課程において想定されている統計の程度よりもやや踏み込んだ内容について扱うのがよいと考える。

生徒たちがこれまで学んできた記述統計は手元のデータを要約するのに対し、推測統計の要は、未知の母集団の様子を標本から推測することにある。そこには確率モデルとして数学的に精緻化された帰納推論（大塚，2020）の思想が組み込まれている。したがって、推測統計を正しく理解するには、その背後にある数学、特に確率に対する理解が必要である。他方、この分野を学ぶにあたり、生徒はまだ積分などの学習を終えていない。そこで、まずは積分と面積の関係に触れることにする。

授業では Python を使用することを踏まえ、面積としての定積分の計算例を、区分求積による近似計算を通して体験する学習を取り入れることにする。Python では統計計算のためのモジュール **scipy.stats** が提供されているため、毎度区分求積によって確率計算を行うわけではないが、ここで一度体験したことが数学Ⅲ（本校では全員履修）の学習で理論的に整備されるというスパイラル学習の効果を狙い、ここで積分の数値計算を敢えて取り入れることにする。

連続型確率分布を学ぶにあたり、理論的な取扱いに深入りせず、直観的に理解できるように（文部科学省，2019）、以下の2点の配慮が必要であると考える。

イ）確率密度関数や累積分布関数について理解を容易にするために、これらのグラフを多用する。

ロ）離散型確率分布で学習した定義や計算方法と対比させる。

特に、ヒストグラムを通じて二項分布と正規分布の関連について考えておくことは、後の標本平均などの性質をシミュレートする際に有効である。

正規分布における確率計算では、確率変数の標準化により、標準正規分布の確率計算に帰着させることが重要である。この計算の手順は、計算機を用いた計算では実は不要である。しかし、後に χ^2 値や t 値を計算する際にも同様の手順を踏む必要がある。この学習においては、生徒は計算機を用いない手計算での確率計算を経験する必要がある。

大数の法則や中心極限定理の取扱いを含め、標本統計の取扱いについては、計算機による乱数生成を用いたシミュレーションとグラフ描画を利用し、直観的に理解することの助けとなるような演習を準備したい。不偏分散と自由度の関係など、理論的な取り扱いが生徒にとって難解であるような内容であっても、数値シミュレーションを行うことで、母分散をよりよく推定する（不偏性）ことが納得できると期待する。

区間推定の数学的な本質は、確率による標準化変数のとり得る範囲から母平均の値を逆算するところにある。推定の授業においては、この点を明示するために、信頼区間の導出を丁寧に行いたい。

区間推定においては、その信頼区間の性質上、一定の確率で母平均を捕らえ損ねる。このことを理解するために、生徒に模擬アンケートを実施し、匿名化した回答結果から生徒が標本を抽出し、得られた信頼区間と母平均を比較するような活動を取り入れる。

仮説検定のアイデアは、生徒にとってすぐに馴染めるものではないだろう。そのため、例となる状況を設定し、仮説検定の思考をたどることを体験しながら、仮説検定の考え方を咀嚼することから授業を始めたい。

仮説検定も、その確率的性質から、誤った検定結果を返す可能性を含む。そのため、検定結果については、その誤りを含む可能性を見積もるべきである。そのための情報として、検出力がある。検出力は、仮説検定に関連する項目の中でも、特に難しいものである。計算例を通じて、その意味を確認しながら生徒が学習を進められるように、図を交えながら丁寧に扱いたい。

生徒の探究の場面、特に総合的な学習・探究の時間における利用を想定し、仮説検定についてはその程度を配慮しつつ、多様な検定手法を提供したい。標準的な z 検定、二項検定や（フィッシャー、ウェルチ） t 検定の他、二群の平均の比較方法について扱うことは、十分に考えられる。また、適合度についての検定などの、 χ^2 検定の豊かな応用例についても触れておきたい。

アメリカ統計学会による p 値の誤用に関する問題提起を皮切りに、近年仮説検定の深刻な誤用が問題視されるようになっている（Wasserstein and Lazar, 2016）。また、仮説検定に代わる新たな統計的推

論・意思決定手法の提案もなされるようになってきている（例えば，Cumming，2014）．初学者が仮説検定を正しく理解するには，乗り越えなければならない障壁は多い．微分積分の学習が十分に進んでいない中等教育段階の生徒であれば，猶更のことである．しかし，統計に限らず，背後の数理についての理解を抜きにして，正しい理解に近づくことは困難であると考ええる．

計算機を用いる際には，パーセント点の扱いや p 値の定義方法に注意すべきである．統計学の理論を解説する教科書等では，パーセント点というと普通は上位から数えた位置を指すのであるが，Excel 等の表計算ソフトでも Python の `scipy.stats` でも，パーセント点は累積分布関数の逆関数として定義されており，下位から数えた位置を表している．また，フィッシャー流の検定における p 値というとき，片側 p 値を表すことも多いが，こちらでも表計算ソフトや Python では両側 p 値のことを指す．これらの差異は，今後生徒がさまざまな資料を参照しながら統計計算を進めていく上で混乱の元となる事項であろう．授業においては，特にパーセント点を用いた計算をよく利用することになるため，この事情については丁寧な取り扱いをしたい．

2 単元の構成

(1) 単元の学習目標

授業や活動を通して，次の事項を身につけることができる．

- (ア) 連続型の確率変数と確率分布について理解し，正規分布などの性質や特徴について理解するとともに，標本調査の考え方について理解を深め，正規分布などを用いた区間推定及び仮説検定の方法を理解することができる．
- (イ) 確率分布や標本分布の特徴を，確率変数の平均，分散，標準偏差などを用いて考察し，目的に応じて標本調査を設計し，収集したデータを基に計算機などの情報機器を用いて処理するなどして，母集団の特徴や傾向を推測し判断するとともに，標本調査の方法や結果を批判的に考察することができる．

(2) 単元の位置

本単元は，「充実期」（3・4 年）における確率・統計分野の最後に位置する内容である．これまでに学んだ確率計算の知識や統計の考え方および収集した統計データの適切な表現手法を基礎として，生徒が，標本から得られる情報を，確率的考察に基づいて整理し，統計的知識に基づいて考察・判断を行い，意思決定に反映させることができるようになることを目指す．

(3) 単元のねらい

- ① 未知の母集団に対して，標本を抽出し，その誤差を適切に見積もり，母集団の情報について調べることができることを学び，また活動を通じて体験することを通して，資質・能力，中でも論理的・批判的・創造的思考の力を育成することをねらいとする．
- ② 資質・能力育成の重点

A 基礎力 (道具や身体を使う)		B 思考力 (深く考える)	C 実践力 (未来を創る)
知識・技能		思考・判断・表現	主体的に学習に取り組む態度
I 知識・概念	II 技能	III 論理的・批判的・創造的思考	IV 自立・協同・創造の力
充実期	連続型確率分布に従う母集団から抽出した標本の性質について理解している．	抽出した標本から母集団の性質について推定したり，仮説を立てて検証したりできる．	抽出した標本から母集団の性質について行った推定や仮説検定の結果について，多角的に考察できる．
			興味のあることに対して小集団で協力してデータをとり，統計的に分析を行い，考察を加えて発表できる．

(4) 単元の展開と評価（全 15 時間）

時	各時の主題	各時の問いと主な活動	各時のねらい（評価の場面）	評価の観点
	事前調査	数学・統計・計算機に関する意識調査		
第 1 次 1 時	積分と数値計算	定積分の値を、区分求積の考えに従い、計算機を利用して計算する。	定積分の近似値を、数値計算によって求めることができる。	Ⅱ
第 2 次 1 時	連続型確率分布	連続型確率分布について、様々な例を通して理解する。	連続型確率分布について、説明することができる。	Ⅰ
2 時	正規分布	正規分布の性質を理解し、正規分布に従う確率変数に関する確率の計算を行う。	正規分布に従う確率変数に対して、所与の区間の確率を求めることができる。	Ⅱ
第 3 次 1 時	標本統計	標本平均や標本分散の分布の性質を理解する。	標本平均の平均や分散について、母集団と関連付けて説明することができる。	Ⅰ
2 時	推定	区間推定の基本的な考え方を理解し、母平均の推定の問題に適用する。	抽出した標本から、母平均の信頼区間を求めることができる。	Ⅲ
3 時	活動	「身の周りにあるものの量を推定しよう」	小集団で協力し、興味をもつ問いを見出し、データを収集し、区間推定を行い、考察し、まとめることができる。	Ⅳ
4 時 5 時	発表	活動内容の発表	小集団でまとめたことを発表し、振り返ることができる。	Ⅳ
第 4 次 1 時	二項検定，t 検定	検定の基本的な考え方を学ぶ。	仮説検定について、説明することができる。	Ⅰ
2 時	対応のある検定，対応のない検定	二群の母平均に差があるかどうかという問題に対する検定手法を学ぶ。	2 種類の母集団から抽出された標本に対し、その母平均についての仮説検定を行うことができる。	Ⅲ
3 時	χ^2 検定	分散の χ^2 検定の手法を学ぶ。また適合度や独立性の χ^2 検定の手法を学び、問題に適用する。	χ^2 検定と応用例について、説明することができる。適合度の χ^2 検定を所与の表に対して適用することができる。	Ⅲ
4 時	活動	「身の周りにあるものについて仮説を立て、検定してみよう」	小集団で協力し、興味をもつ問いを見出し、データを収集し、仮説検定を行い、考察し、まとめることができる。	Ⅳ
5 時 6 時	発表	活動内容の発表	小集団でまとめたことを発表し、振り返ることができる。	Ⅳ
7 時	演習	仮説検定の演習	正しく仮説を立て、仮説検定を行うことができる。	Ⅱ
	事後評価	数学・統計・計算機に関する意識調査		

(5) 評価の観点

Ⅰ 知識・概念 Ⅱ 技能 Ⅲ 論理的・批判的・創造的思考 Ⅳ 自立・協同・創造の力

3 授業記録

(1) 学習過程

Coming soon...

(2) 授業分析

Coming soon...

(3) 成果と課題

Coming soon...

【参考資料】

Wasserstein, Lazar 「The ASA' s statement on p-values: Context, process, and purpose」 American Statistician, Vol.70, No.2, pp.129-133.

Geoff Cumming (2014) 「The New Statistics: Why and How」 Psychological Science, Vol.25, No.1, pp. 7-29.

半田真, 清水克彦 (2021) 「『仮説検定の考え方』の指導に関する実践研究」東京理科大学教職教育研究, Vol.6, pp.33-44.

大塚淳 (2020) 『統計学を哲学する』名古屋大学出版会.

文部科学省 (2019) 『高等学校学習指導要領解説数学編理数編』.

Google1 「Google Workspace」 <https://workspace.google.co.jp/> (2022 年 2 月 8 日閲覧)

Google2 「Colaboratory へようこそ」 <https://colab.research.google.com/> (2022 年 2 月 8 日閲覧)