

中等教育におけるデータサイエンス教育と数式処理

千葉県立千葉中学校・千葉高等学校 大橋 真也

Shinya Ohashi, Chiba prefectural Chiba Junior & Senior High School

1 はじめに

高等学校新学習指導要領におけるデータサイエンスの扱いについて、見本教科書の内容をもとに具体化された内容をもとに考察を行う。また、大学入試共通テストのサンプル問題や大学の数理・データサイエンス・AI教育との接続性について、中等教育におけるデータサイエンス教育として、指導しなければならないことについて検討する。

2 新課程の見本教科書に見るデータサイエンス

新課程の「情報I」の教科書採択のための見本本が、各高等学校に配布されている。高等学校学習指導要領解説情報編の情報Iのデータの分析に関するキーワードについて、教科書にどの程度掲載されているかを手元にある11種類の見本本をもとに調べたものが次の表である。

表 1: 新課程見本教科書におけるデータサイエンスの扱い

教科書	項目 1	項目 2	項目 3	項目 4	項目 5	項目 6	項目 7	項目 8	項目 9
A	○	○	×	×	×	×	×	○	×
B	○	○	×	×	×	○	○	○	○
C	○	○	○	×	×	×	○	○	○
D	○	○	○	×	×	×	○	○	○
E	○	○	×	×	×	○	○	×	×
F	○	○	×	×	×	×	×	○	○
G	○	○	○	×	×	○	×	○	×
H	○	○	○	×	×	×	×	○	×
I	○	○	○	×	×	×	○	○	○
J	○	○	×	○	×	×	×	○	×
K	○	○	○	○	×	×	×	○	×

表の各項目は、次の通りである。

1. 質的データ・量的データ

2. 尺度
3. 異常値
4. データクレンジング, データクリーニング
5. 時系列
6. 交絡因子, 疑似相関
7. 回帰直線
8. テキストマイニング
9. 仮説検定

表を見て分かるが、「時系列」を扱っている教科書はない。高等学校学習指導要領解説では、例示扱いのためだろうか。また、「データクレンジング, データクリーニング」, 「交絡因子, 疑似相関」について扱っている教科書も少なかった。さらに、今回のデータ分析では単回帰分析が必須であると読み取れるが、これさえも扱っていない教科書もあった。これは、高等学校学習指導要領解説と教科書のギャップが大きいということである。このギャップについては、教科会社の編集者や執筆者のデータサイエンスに対する認識不足によるところが多いと考えている。同様のことがプログラミングについても見ることができる。

教科書のデータの分析に使用しているソフトウェアに関しては、表計算ソフトウェアがすべての教科書で使用されており、一部 Python(2種類), JavaScript(1種類), Web ツールなどの仕様もあったが、プログラミングとの融合を考慮している教科書は少なかった。また、R等の統計ソフトウェアや数式処理ソフトウェアを使用している例は全く見られなかった。

教科書に先行して発表された「高等学校情報科『情報 I』教員研修用教材」では、Rなどを使った実習例が多く掲載されており、表計算ソフトウェアで処理することの限界が示されているにもかかわらず、教科書においては表計算ソフトウェアしか使われていないことは残念なことである。

3 高等学校のデータサイエンスとは

以前にも報告したが、高等学校の共通教科情報では、必修科目の「情報 I」と選択科目の「情報 II」が設定されている。「情報 I」のデータサイエンス部分については、前述した教科書調査と重複するので割愛する。「情報 II」の教育課程の中に位置づけている学校は少ないが、以前の専門科目等に比べて、AIなどの知的基盤社会やデータ駆動型社会への認識が高まったことや大学入学共通テストへの「情報 I」の採用により、2, 3年生に設置する学校も見られる。「情報 II」のデータサイエンスの扱いは、教科書がまだ出ていないが、文部科学省から出されている「高等学校情報科『情報 II』教員研修用教材」に整理されているが、次のような内容である。

- データと関係データベース
- 大量のデータの収集と整理、整形
- 重回帰分析とモデルの決定
- 分類による学習
- 主成分分析による次元削減
- クラスタリングによる分類
- ニューラルネットワークとその仕組み
- テキストマイニングと画像認識

4 大学入学共通テストのサンプル問題に見られるデータサイエンス

令和7年度の大学入学試験共通テストから、正式に「情報I」が受験科目に加わった。大学側の採用の対応については、現段階でも未定の部分があるが、この内容に関しては、令和2年12月には、『「情報I」試作問題(検討イメージ)』が大学協会等向けに公開され、令和3年3月には『「情報I」サンプル問題』が一般に公開された。試作問題には、データサイエンスに関する問題はほとんど掲載されていないが、サンプル問題では、データの分析として第3問が公開されている。

第3問 次の文章を読み、後の問い(問14)に答えよ。

世界サッカー選手の一部のサッカーチームを題材としたデータについて調査した。『ワールドカップ』というサッカー大会ではあるが、ここではある特定のサッカー大会について調査している。調査したチーム(対象チーム)と、対象チームのワールドカップにおける、予選で選出したチーム(予選選出チーム)と、予選を通過し、決勝トーナメントに進出したチーム(決勝進出チーム)との違いを、データに基づいて分析することとした。このデータを用いて各項目の各チームの中で、決勝進出チームはサッカー選手、予選選出チームはサッカー選手である。

- 調査数：大会期間中に行った試合数
- 調査項目：大会期間中に試合が行われていた試合の試合数
- サンプルサイズ：大会期間中に試合が行われていた試合の試合数の中から無作為に抽出した試合の試合数
- サンプルサイズ：大会期間中に試合が行われていた試合の試合数の中から無作為に抽出した試合の試合数
- 調査対象：大会期間中に試合が行われていた試合の試合数

調査結果は、決勝進出チームと予選選出チームの両方について、以下の表に示すように、各項目の平均値を算出した。また、チームごとに調査対象となる各項目の1試合あたりの変動係数を算出した。

項目	決勝進出チーム	予選選出チーム	変動係数
試合数	10	20	0.5
ゴール数	15	30	0.6
失点数	10	20	0.5
パス数	100	200	0.5
ドリブル数	50	100	0.5
シュート数	10	20	0.5
キック数	10	20	0.5
パスミス数	10	20	0.5
ドリブルミス数	5	10	0.5
シュートミス数	5	10	0.5
キックミス数	5	10	0.5
パスミス率	0.1	0.2	0.5
ドリブルミス率	0.1	0.2	0.5
シュートミス率	0.1	0.2	0.5
キックミス率	0.1	0.2	0.5

また、アンケートを基に、観客動向データを用いて、図1を作成した。

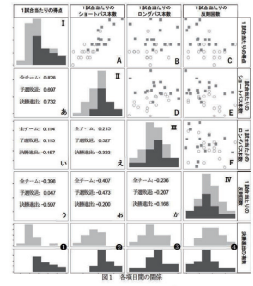


図1 各項目の調査結果
図1の1～12は、それぞれ項目の各チームのヒストグラムと各項目のチームと予選選出チームとの比較したものである。●●●は各項目のチームと予選選出チームとを比較したヒストグラムである。●●●は各項目のチームと予選選出チームとを比較したヒストグラムである。●●●は、それぞれ2つの項目の各チームと予選選出チーム、予選選出チームのそれぞれに選出したチーム数である。またA～Fは、それぞれの2つの項目の各チームと予選選出チームとを比較したヒストグラムと予選選出チームとを比較したヒストグラムである。また、図1のAは観客動向データを用いて作成されたものである。

図 1: 大学入試共通テストサンプル問題

今回公開されたサンプル問題は、ワールドカップのサッカーを題材にしたデータについての問題である。問題を見て顕著な部分は、数式を出さずに図を中心に解答をさせていることである。内容は、散布図・相関行列から、様々な次章を読み解く問題になって

いる。散布図・相関行列は、高等学校学習指導要領解説の中には、「項目間の相関を見るためにデータを漏れのないように組み合わせる複数の散布図などを作成し、相関関係の見られる変数の組合せを見出し、」などとあることから、このような図を作成して読み解かせる問題となっていると考えるが、残念ながら、このような図を掲載している教科書は1冊もなかった。散布図・相関行列は、RやPythonなどを用いれば、数行のプログラムで描画することが可能なグラフであるが、表計算ソフトウェアのみで説明を進めている教科書では、このようなグラフについて考えが及ばなかったのかも知れない。

今後、大学での採用などの情報が広がり、「情報I」に関する共通テストに関する情報も多く出てくると考える。実際に、受験産業や問題集などの関連書籍について、様々な動きが見られはじめています。今後の動向に注目したい。

5 大学における数理・データサイエンス・AI教育

大学では、全学的な「数理・データサイエンス・AI教育」が進められてきている。文部科学省でも初等中等教育、高等教育、そして産業界のエキスパート人材の育成に接続できるように、データサイエンス教育の取り組みを発表している。

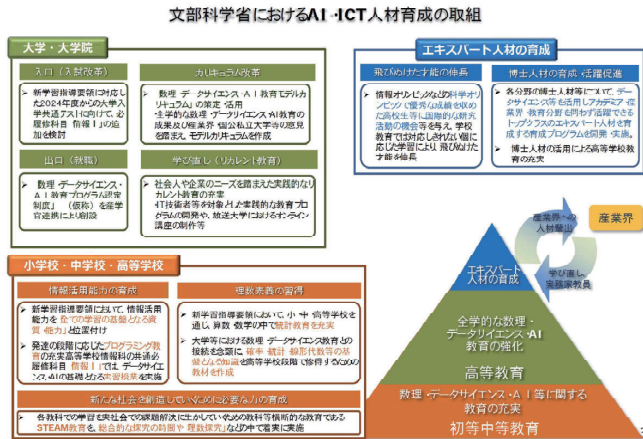


図 2: 文部科学省における AI・ICT 人材育成の取組

高等学校の共通教科の「情報II」では、データサイエンスとして機械学習などを扱っていることから、それを更に理論的にかつ体系的にすべての学生が学習するものと推察される。6大学のコンソーシアムで作成したモデルカリキュラムにもそのような内容を見ることができ、最近公開された東京大学のモデルカリキュラムにおいてもPythonを使った機械学習などの内容が掲載されている。

しかしながら、現実には認定を受けた大学のカリキュラムを見ると、高等学校において数学Bや数学IIIを履修していない、または受験に数学を必要としていない文系学生への配慮であろうか、簡単な記述統計学をデータサイエンスの理論と実習に当てている大

学も多くない。令和7年度に新課程の「情報I」を学んだ生徒が、「情報I」の共通テストを受け、大学で学ぶデータサイエンスがこの程度のものでよいのだろうか。東京大学の提示するモデルカリキュラム程度の内容は必要になると考える。

6 高大のデータサイエンスにおける接続の問題点

このように文部科学省では、データサイエンス教育に関する接続の道筋を作り、提示しているが、それらを実現するためには、様々な問題が残っている。そのひとつが人材不足である。

高等学校においても、新課程の「情報I」や「情報II」を指導できる教員の育成を進めている。前述した「高等学校情報科『情報I』教員研修用教材」「高等学校情報科『情報II』教員研修用教材」を用いて各都道府県の教育センターや情報関連部会によって、研修が進められており、プログラミングやデータサイエンスの実習が行われている。ただし、実際には研修の時間が足りておらず、「情報II」の内容の研修まで行っていき自治体は多くない。また、一方で情報科の教員採用を各都道府県で増やす動きがある一方、全く採用のない県もいまだに存在する。2000年頃から3年間で行った「情報科現職員等教員免許講習」で情報科の免許を取得した先生方も定年退職等により半減しており、さらにこの5年でほとんどが退職の年齢を迎える。この状況で教員不足、データサイエンスを教えられる教員のスキルアップは、大きな問題である。臨時免許や免許なしで「情報I」を教えるのは、これからは難しいことになる。

また他方で、全学的に数理・データサイエンス・AI教育を実施する大学側でもデータサイエンス人材が不足している。データサイエンスを教えられる人材の需要は年々高まる一方、供給が追いついていないのである。それだけ、データサイエンスを教えられる人材を育成してこなかったことが、ここで問題になっている。また、前述した学生の数学などの高等学校の際の履修状況や大学入学試験での教科として採用してこなかったことにより、十分な教育が行えないというのも問題となっている。一部の大学では、文系であっても入学試験で数学を必修化するなどの動きも見られ、今後この状況も変化の兆しが見えてきている。

7 データサイエンスと数式処理ソフトウェア

文系生徒の数学技能の不足や統計ソフトを用いたプログラミングや分析についての問題を解消する1つの方法がある。Courseraのスタンフォード大学で開講している「機械学習」の講座では、Octaveという数式処理ソフトウェアを用いて、数学の部分の実習を行っている。文系学生が数学技能が不足しているならば、数式処理を使ってはどうだろうか。私自身、ある大学でのデータサイエンスの授業において、Mathematicaを用いて、機械学習の実習を行っているが、簡単な命令(関数)で、複雑な機械学習の処理を行うことができるので、学生の理解の助けになっている。

高等学校では、Rさえ使っている教科書が見られないが、せめてRや数式処理を実際に高等学校や大学のデータサイエンスの授業に使ってみてはどうだろうか。

8 おわりに

高等学校のデータサイエンス教育も大学の数理・データサイエンス・AI教育も動き出したばかりである。高大のデータサイエンス教育の接続やそれぞれの教育の質の向上、人材不足などの問題点は、徐々に解消されることを期待したい。

参考文献

- [1] 文部科学省：高等学校学習指導要領,
http://www.mext.go.jp/component/a_menu/education/micro_detail/_icsFiles/afieldfile/2018/07/11/1384661_6_1_2.pdf, 2018.
- [2] 文部科学省：高等学校学習指導要領解説(情報編),
http://www.mext.go.jp/component/a_menu/education/micro_detail/_icsFiles/afieldfile/2018/07/19/1407085_16_01.pdf, 2018.
- [3] 文部科学省：高等学校情報科「情報I」教員研修用教材,
https://www.mext.go.jp/a_menu/shotou/zyouhou/detail/1416756.htm,
2019.
- [4] 文部科学省：高等学校情報科「情報II」教員研修用教材,
https://www.mext.go.jp/a_menu/shotou/zyouhou/detail/mext_00742.html,
2020.
- [5] 文部科学省：数理・データサイエンス・AI教育プログラム認定制度(リテラシーレベル),
https://www.mext.go.jp/a_menu/koutou/suuri_datascience_ai/00002.htm,
2021.
- [6] 数理・データサイエンス教育強化拠点コンソーシアム,
<http://www.mi.u-tokyo.ac.jp/consortium/>.