

# 大学生のためのデータサイエンス 学外実習とアンケート

水 井 雅 彦

## I はじめに

科目名「データサイエンス」が新設され、2年目の講義を迎えた。初年度はコロナ禍への対応にて、オンライン講義での実施となった。受講生の理解や習熟状況の把握は、提出課題やポータルサイト「Q&A」から推察する。しかし、状況把握に限界がある。対面講義が再開され、受講状況や質問から感じ・思うことを、講義改善を目的に整理する。大学内でのデータサイエンス活用を考える。

## II データサイエンス

### 1. データサイエンスとは

平成28年12月文部科学省は「大学の数理・データサイエンス教育強化方策について」\*1の中で、「未来社会においては、広範囲かつ目的に整合したデータの取得を前提として、それらのデータに内在する本質的構造を見極め、数理的思考に基づいて解析・問題解決を行う能力、データサイエンスを活用して新たな価値を生み出し、有用なシステム構築につなげる能力が求められる。」と表現した。これを学生が身に付ける能力として読み替えると、「そこにあるデータを表・グラフなどで視覚化し、統計学の知識から特徴を読み取り、その要因・

原因を察し、問題・課題解決を進める能力を、データサイエンスとする。」と解釈できる。既存の統計学、数学、情報処理分野にまたがる統合型科目と言える。

情報通信技術（Information and Communication Technology）の普及から、データを豊富に入手・管理できる基盤は整う。しかし、データに基づく状況解析からの意思決定を行う企業割合が低く、GE グローバル・イノベーション・バロメータの2016年調査結果では世界平均61%に対し、日本は40%にとどまる。また、国内の98.4%が中小企業である点からも、データ活用の専門家を雇えない、育成出来ない、現状業務の変更を考えていない、などが多数派であると考えられる。生産性向上や新産業創出の強化に、データ活用の意義は大きい。そこで、大学でのデータを活用人材育成が求められる。

大学での教育では、専任教員の配置や、数理系科目とデータサイエンスにて数理的思考やデータ分析・活用能力をバランスよく習得させる教育内容など、具体的な方針が示される。本学では、一般教養にて開講する。標準カリキュラムや教材の内容への改善に対し、近隣大学間でのコンソーシアム設立などにも言及されている。

## 2. 数理・データサイエンス・AI 教育プログラム認定制度

文部科学省高等教育局は、「数理・データサイエンス・AI 教育プログラム認定制度」\*<sup>2</sup>を設け、ICT 活用を推進する大学の認定を行う。本件のリテラシーレベル講義を1年以上実施し、全ての学生が履修できるカリキュラムであることや、数理・情報系科目の連携が取れていることなどが条件となる。

認定に向けたカリキュラム更新は、明確な目標に沿って実施されるため、大学が社会へ輩出する人材像を具体化できる利点がある。本学の強みである学外実習では、ボランティアや地域住民との課題解決などの実績がある。この活動実績の評価をキーワードとして、活動 → アンケート調査 → データ分析 → 活動の改善検討 → 改善内容の実施 → 再評価 のサイクルを回すための、デー

タサイエンス講義内容を計画する。

### Ⅲ 対面講義

#### 1. 数学への苦手意識

高等学校の数学Ⅰでは「データの分析」を章で扱い、四分位数と箱ひげ図、分散と標準偏差・相関関係について学ぶ。また、数学Aでは「場合の数と確率」を章で扱い、集合と順列・組み合わせ、場合の数と確率について学ぶ。この学習内容を基に、サンプル数の少ないアンケート結果の把握などは、十分に実施可能である。しかし、昨年実施したオンライン講義は、ポータルサイト「Q&A」や対面での質問対応で、「初めて意味が判った」などの感想が多く、高校での学習内容が活用できていない事が推測される。

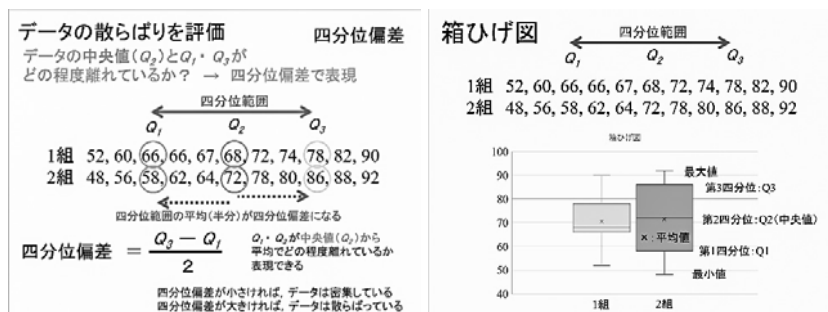


図1. 統計学の再学習

そこで、受講生が該当の教育内容を始めて学ぶ想定で講義を進めた。対面講義では、教材提示や板書の内容をノートに纏める熱心さから、講義内容へ興味を示しているか推察できる。受講生が数学に対する苦手意識を感じると、ノート記入に対する作業時間が伸びる点を確認できる。この場合、講義内容の活用事例や利便性を説明し、少しでも講義内容に興味関心を持てるように時間を取っている。

## 2. 推理の楽しさ

図2に、共に平均65点の、テストA・テストBの得点を示す。

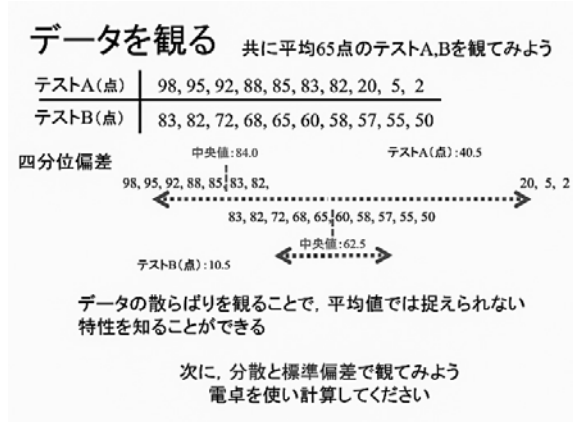


図2. 平均が65点のテスト

受講生へ2つのテストに共通する82点に対し、どちらの82点が評価されるべきか？を推理する課題を求めた。四分位数（箱ひげ図）と標準偏差から、データの「バラツキ」を検討すると、テストAの20点、5点、2点が、外れ値であることが確認できる。外れ値の発生原因とする理由を推理させると、「この3名は、勉強していなかったのでは？」と考えられる。そこで、2つテスト

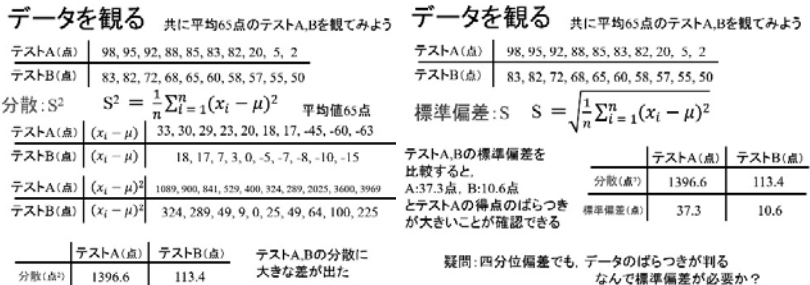
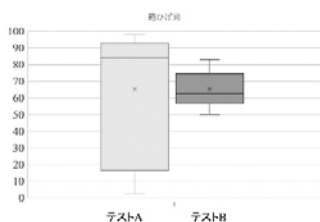


図3. 分散と標準偏差

の得点を違った視点から比較するため、図4に外れ値を除外した箱ひげ図を作成した。これによりデータの比較を行えば、視覚的にテストAの平均点上昇とデータのバラツキを確認できる。サンプル数の少ないアンケート結果などの把握では、四分位数によるデータのバラツキの把握を薦める。

平均は同じだが・・・



回答例

提出課題：  
平均値65点のテストA・Bで、どちらの82点が頑張ったか？

テストA(点)	98, 95, 92, 88, 85, 83, 82
テストB(点)	83, 82, 72, 68, 65, 60, 58, 57, 55, 50

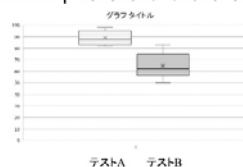


図4. 外れ値を除外して再評価

講義では受講生の自由な推理を促し、様々な視点からデータを推理する楽しさを伝える。外れ値もデータであるが、これを除外すれば結果は大きく変化する。数学の知識を道具として使う小さな推理を講義に取り入れ、苦手意識の克服を計る。

### 3. 表計算ソフトの活用

オンライン講義でも活用したマイクロソフト社 オフィス360A1は、表計算ソフト：エクセル上で、ヒストグラム、散布図や箱ひげ図など様々なグラフ作成が可能である。講義ではグラフによる視覚効果を説明し、表現したい目標に合ったグラフの選択を扱う。

また、データの並び替え：ソートや、データの条件適合数カウント：COUNTIF 関数の実演を行い、手作業と便利な機能の比較を行う。

## IV アンケート調査

マイクロソフト社 フォームスは、web ブラウザ上で動作するアンケートや試験などの設問、解答欄、選択肢などの書式作成と、解答結果をエクセル上で分析・評価できるアプリケーションである。

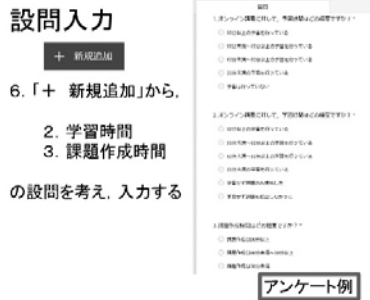


図5. アンケート例

これを10インチ程度の画面サイズを持つタブレット端末で運用すれば、屋外でも容易にアンケート調査が実施できる。データサイエンスでは、現実社会でのデータ取得と、その評価が重要である。本学の学外実習での活動状況評価を想定し、講義内容を計画した。

アンケートの質問や選択肢の構築には、「問題解決」\*3の手法を用いる。明らかにしたい問題を小さな問題に分解し、その原因を明らかにすれば、大きな問題に対応できる。アンケートの質問。選択肢は、分解された小さな問題に対する確認として構築する。

## V おわりに

データサイエンスの講義では、数を用いて状況判断を行う小さな体験を重ね、「これ便利だ」と感じてもらう講義内容に努めている。一般教養の講義内容と

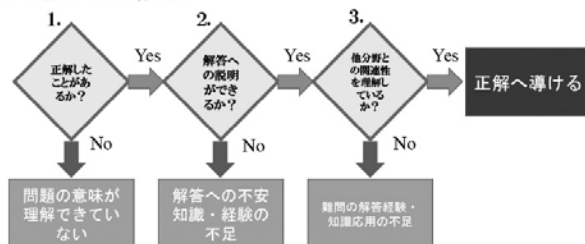
## ②原因の究明

算数が苦手な小学生  
(6年生)を想定してみよう  
→目的:算数への学習対応を考える

### 1. 原因としてあり得るモノを洗い出す

木分解の内容を Yes, No の質問で置き換える

#### 3つの質問で再構築



関係性を図式化して整理

図6. 原因の究明

なるが、本学の強みである学外実習の活動評価にて、アンケートを中心とした評価が行われれば、教育内容の定着に繋がる。対面講義の15回を終えたのち、再び内容検討を行いたい。

## 参考文献

1. 文部科学省, 「大学の数理・データサイエンス教育強化方策について」, 平成28年12月12日, 2018. [https://www.mext.go.jp/b\\_menu/shingi/chousa/koutou/080/gaiyou/1380788.htm](https://www.mext.go.jp/b_menu/shingi/chousa/koutou/080/gaiyou/1380788.htm)
2. 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)」, 令和4年3月15日, 2022. [https://www.mext.go.jp/a\\_menu/koutou/suuri\\_datascience\\_ai/00002.htm](https://www.mext.go.jp/a_menu/koutou/suuri_datascience_ai/00002.htm)
3. 水井雅彦, 「データサイエンス教育」教養研究, 九州国際大学, 第28巻, 第3号, pp. 21-31.