

2022年度「データサイエンス基礎」

初回アンケート集計結果

1. 概要

本アンケートは、履修者のデータサイエンスに対する認識や期待、高校での数学や統計の履修状況、パソコン操作の習得状況を把握し、今後のデータサイエンス教育の改善と自己点検評価に役立てることを目的として授業の第1回目に実施した。最終的に回答者は1368人（説明を読み、「同意しない」を選択した学生が1名）となり、学部ごとの回答状況を表1に示す。

表1. アンケートの学部ごとの回答状況

学部	履修者数	回答者数	回答率
医学部	325	324	0.997
教育学部	170	168	0.998
人文社会科学部	292	280	0.959
農学生命科学部	239	227	0.950
理工学部	374	368	0.984

2. 高校数学でのデータサイエンスに関連する内容の履修状況

履修者に「数学I」、「数学A」と「数学B」のデータサイエンスに関連する内容の履修状況について、4段階の主観評価（全く知らない・履修していない、名前は聞いたことがあるがよくわからない、ある程度自分で理解している、他の人に説明できる）を求めた。その結果を図1に示す。ほとんどの学生が「数学I」の「データの分析」と「数学A」の「場合の数と確率」を履修しているのに対し、「数学B」の「統計的推測」の履修の割合は低くなっている。この傾向は学部間にも大きな差が見られなかった。

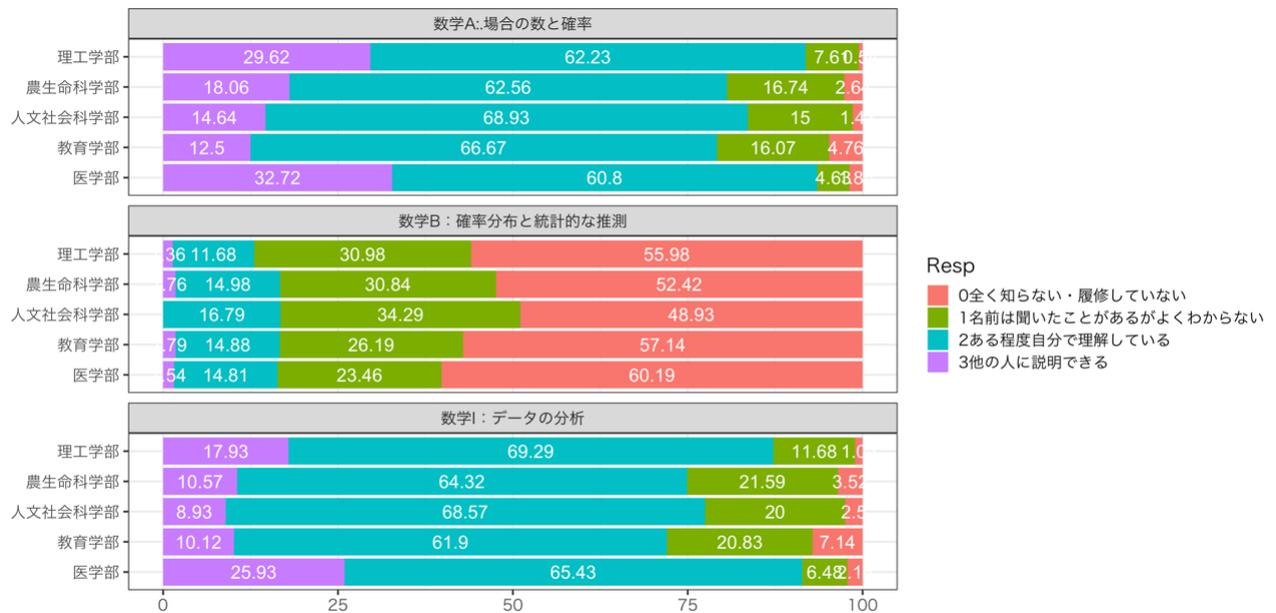


図1.学部ごとの高校数学でのデータサイエンス関連分野履修状況分布

3. 受講前のデータサイエンスの認知度、PCスキルについて

(ア) データサイエンスの認知状況 (問5)

履修者には「データサイエンス」(アンケートの設問で「データサイエンスの説明を与えている」)をどの程度知っているかを4段階評価(全く知らない・履修していない～完全に理解している)を求めた。その結果は図2に示した。ごく一部の学生しか理解していないことが確認できた。

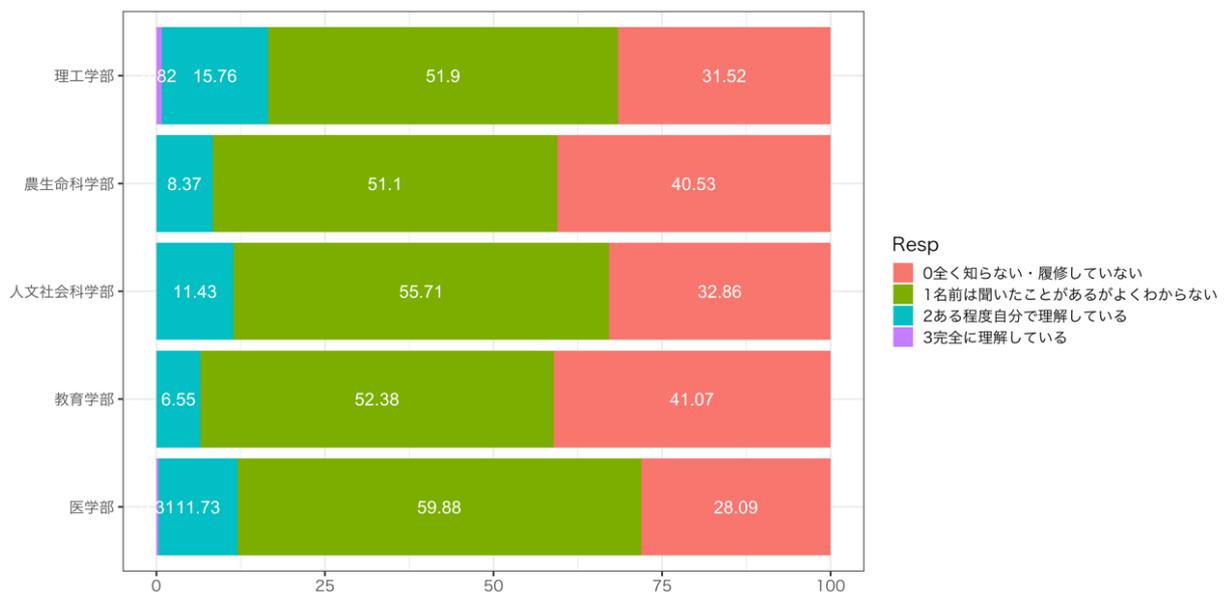


図2.学部ごとのデータサイエンスについての熟知状況

(イ) 確率・統計とデータサイエンスの用語の認知状況 (問4、6)

履修者に確率・統計と関連する言葉をどの程度知っているかについて2択での回答(Yes/No)を求めた。その結果は図3に示した。中学数学で学ぶ「平均」「分散」などの言葉は多くの人知っている。一方、統計分析の基礎である「確率変数」、「標本」、「分布」(正規・二項分布)と「回帰分析」についてはほとんど知らないことがわかる。



図3.統計に関連する言葉の認知状況

さらに、データサイエンスに関連する言葉をどの程度知っているかについても2択で回答を求めた。その結果は図4に示した。近年の話題である「人工知能」と「SDGs」の認知度は高いが、「ELSI」や「データ駆動型社会」などの認知度はかなり低い。



図 4.データサイエンスに関連する言葉の認知状況

(ウ)PC の操作に関する調査 (問7)

履修者にパソコンのキーボード操作について 2 択評価 (Yes/No) を求めた。その結果は図 5 に示した。ほとんどの学生はキーボードの使用経験があるが、タッチタイピングに関してはできる人とできない人が半々に分かれている。

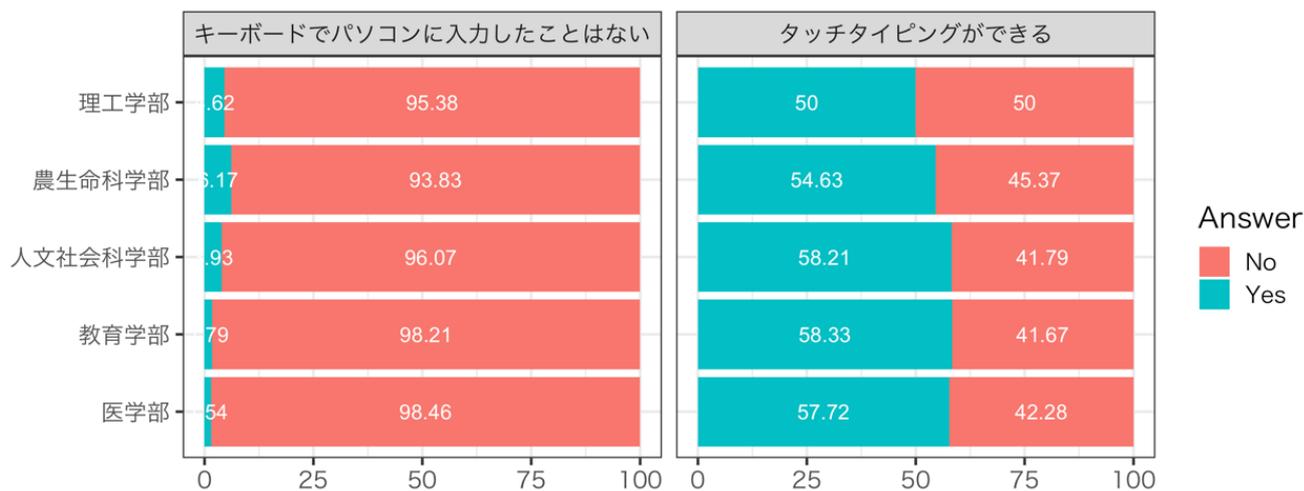


図 5.学部ごとの PC 操作についての状況分布

(エ) プログラミング言語の学習状況 (問8)

履修者にプログラミング言語の学習状況についても 2 択で回答を求めた。その結果は図 6 に示した。ほとんどの学生はプログラミングの未経験者であることが確認できた。



図 6.プログラミング言語の学習状況

(オ) Office の使用経験 (問9 - 10)

履修者にはマイクロソフト社のソフトを学んだこと・使ったことがあるかどうかについて 2 択で回答求めた。その結果は図 7 に示した。ほとんどの学生はオフィスのメインソフトを使用したことがあると回答した。

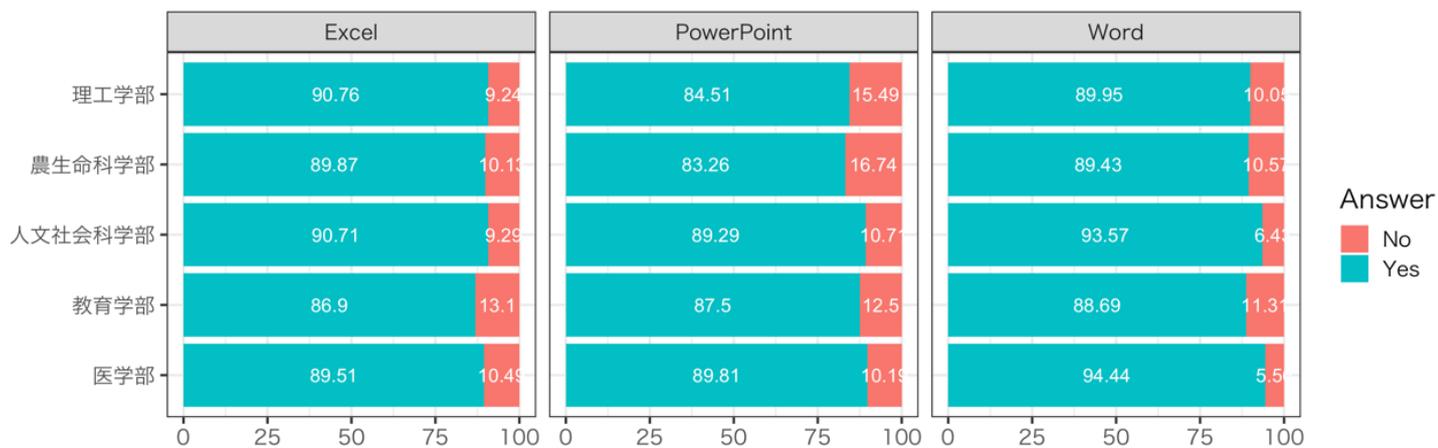


図 7.オフィスの使用状況

さらに、授業で使用する Excel の修得状況についても確認した。その結果は図 8 に示した。棒グラフの作成や平均値の計算に関して一部のできる学生を除き、ほとんどの操作ができないという状況が確認できた。

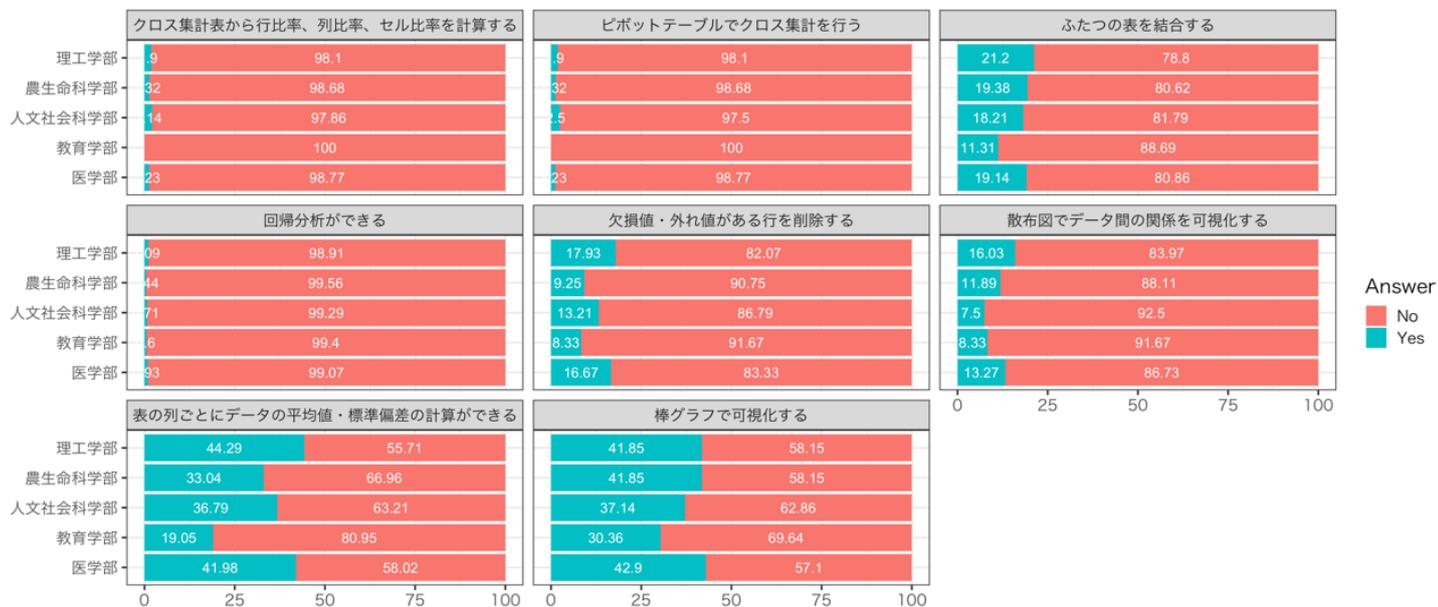


図 8.Excel の使用状況

4. まとめ

本アンケートの結果から、履修者は数学や統計に関する知識がある程度持っているものの、データサイエンスに関する認識と理解がまだ浅く、基礎的な教育が必要であることが確認できた。活用の面においては Excel の使用状況や熟練度も低く、演習を兼ねた教育の重要性を示唆している。

Appendix. 「データサイエンス基礎」初回アンケート

説明

本アンケートの目的は、新入生の皆さんのデータサイエンスに対する認識や期待、高校での数学や統計の履修状況、パソコン操作の習得状況を把握し、今後のデータサイエンス教育の改善に役立てることです。

回答は Forms から行い、Forms の機能により匿名のままデータ化されます。さらに、収集されたデータを上記の目的に沿って分析し、教育資料や研究発表などの場において公表・公開する際には、集計データのみを公開・公表とします。あなたの答えが個人として分析されることは絶対にありませんし、お答えにより何らかの不利益が生じることもありません。この点をあらかじめご理解とご了承をお願いします。

次の URL からアンケートの回答画面に進んでください。以上の説明でも回答に同意されない場合、最初の設問で「同意しない」を選んでください。アンケートはその時点で終了します。お忙しいところ恐れ入りますが、ぜひとも答えていただくよう、お願いいたします。

設問は9問です。個人差はありますが 回答には 10 分程度かかります。よろしくお願ひします。

Part 1. はじめに、あなた自身の所属などについてお尋ねします。(1問)

・所属学部科を選んでください。

1. 人社・人文学部
2. 教育学部
3. 医学部医学科
4. 医学部保健学科
5. 理工学部
6. 農学生命科学部

Part 2. 「数学」の履修状況について質問します(2問)

・次の「数学」の項目について理解度を 0 – 3 点で答えて下さい。主観的な判断で構いません。(0点=全く知らない・履修していない、1点=名前は聞いたことがあるがよくわからない、2点=ある程度自分で理解している、3点=他の人に説明できる)

1. 数学 I：データの分析
2. 数学 A：場合の数と確率
3. 数学 B：確率分布と統計的な推測

・次の言葉（または説明文）を知っていますか？ Yes か No でお答えください。

1. 平均値
2. 最頻値
3. 中央値
4. 四分位範囲
5. 分散
6. 標準偏差
7. 箱ひげ図
8. ヒストグラム
9. 散布図
10. 相関関係
11. 共分散
12. 相関係数
13. 相関関係と因果関係は等しくない
14. 確率変数
15. 期待値
16. 二項分布
17. 正規分布
18. 標本平均、標本比率
19. 統計的推定
20. 回帰分析

Part 3. パソコン操作やデータサイエンスについて質問します(6問)

データサイエンスとはデータを用いて学術的・社会的な課題解決を行う新しい学問分野です。数学、情報科学などの学問分野と解決すべき課題が属する分野の知識（ドメイン知識）に基づいてデータを分析し解決する方法論を扱います。以下、データサイエンスとパソコン操作、プログラミングについて質問します。

・あなたはデータサイエンスをどの程度知っていると思いますか？0－3点で答えて下さい。主観的なもので構いません。（0点＝全く知らない・履修していない、3点＝完全に理解している）

・次の言葉を知っていますか？Yes か No でお答えください。

1. 人工知能 (A.I.)
2. 機械学習
3. データ駆動型社会
4. ビッグデータ

5. ディープラーニング
6. ELSI
7. 統計的差別
8. SDGs

・パソコンのキーボード操作についてお伺いします。Yes か No でお答えください。

1. タッチタイピングができる。
2. キーボードでパソコンに入力したことはない。

・次のプログラミング言語を学んだことがありますか？Yes か No でお答えください。

1. Python
2. JavaScript
3. C
4. その他の言語

・次のマイクロソフト社のソフトを学んだこと・使ったことがあるかどうかを Yes か No でお答えください。

1. Excel
2. Word
3. PowerPoint

・Excel を用いて次の操作が出来ますか？Yes か No でお答えください。

1. 表の列ごとにデータの平均値・標準偏差の計算ができる
2. 欠損値・外れ値がある行を削除する
3. ふたつの表を結合する
4. 棒グラフで可視化する
5. ピボットテーブルでクロス集計を行う
6. クロス集計表から行比率、列比率、セル比率を計算する
7. 散布図でデータ間の関係を可視化する
8. 回帰分析ができる