

「大学における数理・データサイエンス教育の全国展開」

大学におけるデータサイエンス教育に関する スキルセット及び学修目標

第2次報告

令和3年（2021年）6月

数理・データサイエンス教育強化拠点コンソーシアム

カリキュラム分科会

要 旨

1 作成の背景と経緯

- 数理・データサイエンス教育強化拠点コンソーシアムでは、「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 数理及びデータサイエンス教育の強化に関する懇談会）を受け、文理を問わず、全国的なモデルとなる標準カリキュラムを検討。これまでに以下を公表。
 - 2019 年 11 月「データサイエンス教育に関するスキルセット及び学修目標 第 1 次報告」（カリキュラム分科会）
 - 2020 年 4 月「数理・データサイエンス・AI（リテラシーレベル）モデルカリキュラム ～ データ思考の涵養 ～」
 - 2021 年 3 月「数理・データサイエンス・AI（応用基礎レベル）モデルカリキュラム ～ AI×データ活用の実践 ～」
- 本報告は、2019 年 11 月に公表したカリキュラム分科会報告の第 2 次報告として、主に中級レベル、そしてさらに上級レベルへの橋渡しを含むスキルセット及び学修目標を整理したもの。

2 スキルセット及び学修目標（第 2 次報告）の性格と役割

- 本報告は、第 1 次報告を包含しつつ、主に中級レベル、そしてさらに上級レベルへの橋渡しとなるべく、より高次のスキルセットを網羅的に示したもの。
- 第 1 次報告と同様に、各大学が一律に整備することを意図したものではなく、各大学の教育目的、分野の特性、個々の学生の学習歴等を考慮し、必要に応じて上位互換や既習確認等を行いながら、適切かつ柔軟に選択・抽出されることを想定。
- 本報告を活用し、いくつかの大学の協力を得て実践モデルの試行、グッドプラクティスの紹介等に取り組むことで、データサイエンス教育の全国への普及・展開を促進。

3 各大学におけるカリキュラムの実装にあたっての期待

- 第 1 次報告では、初級レベルの教育に当たり、「なぜ数理・データサイエンス・AI を学ぶのか」、「社会でどのように活用されているのか」など、学修の動機付けが重要であることを提示。応用基礎レベル以降の教育では、データから意味を抽出し、現場にフィードバックする能力、AI を活用し課題解決につなげる基礎的な能力を持ち、様々な専門分野で活用・応用展開しながら現実の課題解決、価値創造を担う人材を幅広く育成することが必要。
- その際、実データ、実課題を用いた演習など、社会での実例を題材とし、現実の課題へのアプローチ方法及び数理・データサイエンス・AI の適切な活用法を学ぶことを組み入れたり、各専門分野の特性に応じた演習や PBL 等を効果的に組み入れたりすることにより、実践的スキルの習得を目指すことを期待。

4 スキルセット及び学修目標（第2次報告）の構成

- 第1次報告との互換性を保つため、「データサイエンスを学ぶ意義」、「データの法規・倫理」、「データの記述・可視化」、「データの取得・管理・加工」、「統計基礎」、「数学基礎」、「計算基礎」、「モデリングと評価」の大分類、その下に中分類・小分類を置く階層構造。
- 小分類を単位として、スキルセットと学修目標を整理。スキルセットは、データサイエンスの習得すべきスキル・理解すべき概念のセット。学修目標は、スキルや概念のイメージを文章でまとめたもの。
- 大分類等に対応させて科目設計を行うことは必ずしも意図していない。スキルセットと学修目標を参照し、各大学の教育の特徴を活かしつつ、教育資源等も考慮しながら、各大学において、科目編成やシラバス作成等を行うことを想定。

大分類・中分類・小分類の構成

大分類	中分類	小分類	レベル
データサイエンスを学ぶ意義	(データサイエンスを学ぶ意義, 重要性の理解) データ駆動型社会 データに基づく課題解決, 意思決定の支援 データサイエンスのサイクル データサイエンスの様々な事例		★
データの法規・倫理	情報倫理, 情報セキュリティ	情報倫理・関連法規	★
		情報セキュリティ	★
	データに関連する法律・規制	個人のデータに関連する法規	★
		著作権法	★
		統計法	★
		不正競争防止法	★★
		医療ビッグデータ法	★★
	データサイエンスの倫理	倫理に配慮したデータ収集	★
		倫理に配慮したデータ利活用	★
		データの匿名化	★
		データサイエンスに関する様々なバイアス	★
		公表バイアス	★
データの記述・可視化	データの記述	種々のデータ	★, ★★
		基本統計量	★
		相関関係	★, ★★
	データの可視化	グラフの構成要素	★
		統計グラフ	★
		分布の統計グラフ	★, ★★
		ダッシュボード	★★
		グラフィカルモデル	★★★★
データの取得・管理・加工	データ取得とオープンデータ	日本や世界のオープンデータ	★
		オープンデータの取得	★
	データ管理とデータ形式	表形式のデータ	★
		その他のデータ形式	★
		データベース	★, ★★
	データの前処理	データクレンジング	★
		外れ値, 異常値, 欠損値	★
		データ加工	★
統計基礎	確率と確率分布	場合の数, 順列・組み合わせ	★
		確率	★
		確率分布, 確率変数	★, ★★
		主要な確率分布	★, ★★
		確率変数の漸近的性質	★★
	データ収集法と確率構造	標本調査	★
		誤差の扱い	★
		ランダム化比較試験	★
	推測統計	統計的モデル	★
		標本分布	★, ★★
		点推定, 区間推定	★, ★★, ★★★★★
		仮説検定の概念	★
		汎用的な検定	★★
		種々の検定	★★
		多重比較	★★★★
		ベイズ統計的推測	★★★★
	ベイズ理論	事後確率	★
		事前分布・事後分布	★★
		ベイズ的仮説検定	★★
		分布の推定	★★★★
	計算統計	ブートストラップ	★★★★
		疑似乱数	★★★★
		サンプリング	★★★★
		モンテカルロ積分	★★
数学基礎	線形代数	ベクトル	★
		行列	★, ★★
		データ記述と線形代数	★★
		固有値と固有ベクトル	★★, ★★★★★
		n 次元ユークリッド空間	★★, ★★★★★
		数値計算と線形代数	★★, ★★★★★
	微積分	多項式関数	★

		初等関数	★, ★★
		1 変数関数の微分法	★, ★★
		1 変数関数の積分法	★, ★★
		2 変数関数の微分法	★, ★★, ★★★★★
		2 変数関数の積分法	★, ★★, ★★★★★
		数値積分	★★, ★★★★★
	数列	データと数列	★
		数列	★, ★★
	演習	線形代数演習	★, ★★
		微積分演習	★, ★★
計算基礎	情報, コンピュータの仕組み	数と表現	★
		有効数字, 計算の誤差	★
		データ量の単位	★
		デジタル化	★
		文字の表現	★
		集合と命題	★
		論理演算	★
	データ構造	配列, リスト	★
		データフレーム	★
		連想配列	★
		スタック, キュー	★★
		グラフ, 木構造	★★★★
	アルゴリズムとプログラミング	アルゴリズム	★, ★★, ★★★★★
		プログラミング	★, ★★, ★★★★★
モデリングと評価	モデリングによる課題解決	モデル作成	★
	教師あり学習	回帰分析	★, ★★
		判別分析	★★, ★★★★★
		正則化法とモデル選択	★★
		非線形モデリング	★★★★
		決定木	★★
		ベイズ統計・モデリング	★★★★
		強化学習	★★★★
	教師なし学習	クラスタリング	★
		主成分分析	★★
		多様体学習	★★★★
		カーネル密度推定	★
	時系列データ	時系列モデリング	★★, ★★★★★
	生存時間解析	生存時間データ	★★
		生存関数の推定	★★, ★★★★★
		生存時間データの比較	★★★★
	質的データ解析	質的データの解析	★★★★
	テキストデータ解析	テキストデータの数値化	★★★★
		テキスト分析	★★★★
	モデルの評価	モデル評価指標	★, ★★
		訓練データとテストデータ	★
	因果推論	因果モデル	★★
		グラフィカルモデリング	★★, ★★★★★
	深層学習・ニューラルネットワーク	ニューラルネットワークの仕組み	★★, ★★★★★
		深層生成モデル	★★★★

まえがき

情報通信・計測技術の飛躍的發展によって、従来とは質・量ともに全く異なるビッグデータが得られるようになった。ビッグデータの活用領域は予測、意思決定、異常検出、自動化、最適化など多岐に亘り急速に拡大しており、自動運転、画像認識、医療診断、防犯、コンピュータゲームなどデータの活用が従来の社会常識を大きく変えつつある例は枚挙に暇がない。企業の株価総額ランキングの急激な変化が示唆するように、経済成長の要因も従来の労働力・資本・技術革新からデータやデータから価値を生み出す技術へ大きくシフトしている。その結果、数理・データサイエンス・AIなどのサイバー世界の情報処理技術が今後の社会発展の鍵となっている。

このようなデータ駆動型社会への転換を推進していくためには、データから価値を創出できる専門人材であるデータサイエンティストの育成が不可欠であるが、現状では圧倒的に不足しており、世界的にデータサイエンティストの育成が急速に進められている。専門人材の育成と同時に国民全体のリテラシーレベルの向上も不可欠である。我が国の近代化実現において国民の高いリテラシーが果たした役割はしばしば指摘されているが、データ駆動型社会への転換においては、データリテラシーをすべての国民が備えるべき新たな素養としてその涵養に努める必要がある。

しかしながら、急速にデータサイエンスの教育体制整備が行われた欧米先進国と異なり、我が国ではその教育体制の整備が著しく立ち遅れている状況が続いてきた。このような状況を早急に解消しデータサイエンス教育の普及を目指して、2016年度に6大学が数理・データサイエンス教育強化拠点校として文部科学省に選定され、それぞれの大学において全学データサイエンス教育の普及に向けた取り組みを実施してきた。この取り組みの重要な点は、単に自大学での数理・データサイエンス教育の充実を実現するだけでなく、分野を問わず全ての学生を想定して、学生が習得すべき標準的な内容を示したカリキュラム、教材、教育用のデータベースの整備を目指していることにある。そのために、6拠点校はコンソーシアムを形成して、3つの分科会のもとで全国の大学へのデータサイエンス教育の普及に向けた活動を行ってきた。2019年度には新たに20大学が協力校として加わり、更に2020年度以降は公私立大学・短大・高専が連携校として参画し、約110大学等（2021年3月時点）に規模を拡大して数理・データサイエンス・AI教育の普及を加速する活動を展開している。

今回の報告書は、2019年11年に公表した第1次報告（分野を問わず全国の全ての大学生・高専生を想定した初級レベルのスキルセットと学修目標）に引き続き、主に中級レベル、そしてさらに上級レベルへの橋渡しとなるべく、より高次のスキルセット及び学修目標を整理したものである。コンソーシアムでは、本分科会報告を活用しつつ、引き続き全国規模の大学間ネットワークの活性化、産業界、関係省庁、地方自治体等との連携強化を図り、我が国の未来を担う人材の育成を支援・促進していく。

数理・データサイエンス教育強化拠点コンソーシアム カリキュラム分科会

目 次

まえがき

I	スキルセット及び学修目標（第2次報告）について	1
1	スキルセット及び学修目標の位置づけ	1
2	第2次報告の構成	2
3	参考資料等	2
II	データサイエンス	3
1	データサイエンス教育	3
2	データサイエンスのサイクルと重要な分野	4
3	カリキュラム分科会の取り組み	6
4	スキルセットと学修目標	10
5	高大接続	11
III	スキルセット	13
IV	学修目標	23
	データサイエンスを学ぶ意義	25
	データの法規・倫理	27
	データの記述・可視化	31
	データの取得・管理・加工	34
	統計基礎	37
	数学基礎	42
	計算基礎	49
	モデリングと評価	52
附属資料1	高等学校学習指導要領とスキルセット及び学修目標の対応	59
附属資料2	第1次報告のスキルセット 初級レベル（コア☆のみ）	61
附属資料3	第1次報告のスキルセット 初級レベル（☆及び★）	63

（参考） 数理・データサイエンス教育強化拠点コンソーシアムの概要

I スキルセット及び学修目標（第2次報告）について

1 スキルセット及び学修目標の位置づけ

数理・データサイエンス教育強化拠点コンソーシアムでは、「大学の数理・データサイエンス教育強化方策について」（2016年12月 数理及びデータサイエンス教育の強化に関する懇談会）を受け、コンソーシアム内に設置したカリキュラム分科会において、文理を問わず、全国的なモデルとなる標準カリキュラムの検討を進めてきた。

2019年11月には、リテラシーレベルのモデルカリキュラム検討の基礎とする観点から、リテラシーレベルのスキルセット及び学修目標を整理し、第1次報告を公表した。これは「AI戦略2019」に掲げられた具体目標「文理を問わず、全ての大学・高専生（約50万人卒/年）が、課程にて初級レベルの数理・データサイエンス・AIを習得」への対応を見据えた取り組みの一環でもあり、2020年4月にコンソーシアムが公表した「数理・データサイエンス・AI（リテラシーレベル）モデルカリキュラム～データ思考の涵養～」と相補的に用いることで、習得すべきスキルや理解すべき概念を共有し、数理・データサイエンス・AI教育の更なる強化・充実に資することを企図したものである。

本報告は、第1次報告を包含しつつ、主に中級レベル、そしてさらに上級レベルへの橋渡しとして、それと同時に「AI戦略2019」に掲げられた「応用基礎レベル25万人/年」から「エキスパート2,000人/年」への橋渡しとなるべく、より高次のスキルセットを網羅的に示したものである。第1次報告と同様に、各大学・高専（以下「大学等」という。）が一律に整備することを意図したものではなく、各大学等の教育目的、分野の特性、個々の学生の学習歴等を考慮し、必要に応じて上位互換や既習確認等を行いながら、適切かつ柔軟に選択・抽出されることを想定している。

コンソーシアムでは、スキルセット及び学修目標やこれに続くモデルカリキュラムの検討と並行して、放送大学やMOOC（Massive Open Online Course）との連携や、eラーニング教材、教科書及び教育用データベース等の活用を見据えた検討を進めており、必要に応じてこれらの教育資源を併用し教育内容を補完するなど、各大学等の主体的な取り組みにより、データサイエンス教育の推進・強化が図られることを期待している。

今後、本報告を活用しつつ、産業界や国公立大学等関係者の参画を得て実践モデルの試行、グッドプラクティスの紹介等に取り組むことで、データサイエンス教育の全国への普及・展開を図る。

なお、今回報告するスキルセット及び学修目標は、現在在籍する学生を視野に作成している。高大接続の観点から「情報I」の必修化等の高等学校学習指導要領改訂を踏まえ、見直すことを前提としている¹。

数理・データサイエンス・AI分野の発展は日進月歩であり、ビジネスや社会との関わりが急速に深化している。こうした状況に鑑み、様々な社会的課題の解決に向けて、産業界や国・地方公共団体等との連携を図り、実データ、実課題を用いた演習など、社会での実例を題材とし、現実の課題へのアプローチ方法及び数理・データサイエンス・AIの適切な活用法を学ぶことを組み入れたり、各専門分野の特性に応じた演習やPBL等を効果的に組み入れたりすることにより、実践的スキルの習得を目指すことが期待される。第1次報告では、初級レベルの教育に当たり、データサイエンスを専門としない学生に対しても「なぜ数理・データサイエンス・AIを学ぶのか」、「社会でどのように活用されているのか」など、学修の動機付けに留意が必要であることを示した。中級レベル以降の教育に

¹ 高等学校学習指導要領解説とスキルセット及び学修目標との対応については、附属資料1を参照のこと。

当たっては、データから意味を抽出し、現場にフィードバックする能力、AI を活用し課題解決につなげる基礎的な能力を持ち、様々な専門分野で活用・応用展開しながら現実の課題解決、価値創造を担う人材を幅広く育成することが重要であろう。各大学等・教員の創意工夫によって Society 5.0、AI 時代に対応した人材育成を強化・推進することを期待している。

2 第2次報告の構成

第2次報告は、初級レベルを包含しつつ、主に中級レベル、そしてさらに上級レベルへの橋渡しを含むスキルセット及び学修目標からなる。スキルセットは、データサイエンスの習得すべきスキル・理解すべき概念のセットである。学修目標は、スキルや概念のイメージ、習得すべき内容を文章でまとめたものである。それぞれ、「データサイエンスを学ぶ意義」、「データの法規・倫理」、「データの記述・可視化」、「データの取得・管理・加工」、「統計基礎」、「数学基礎」、「計算基礎」、「モデリングと評価」の分野（大分類）で構成している。

なお、分野（大分類）等に順序性はなく、各大学等の実情に応じてカリキュラム編成等が行われることを想定している。

3 参考資料等

カリキュラム分科会では、スキルセット及び学修目標の検討に当たり下記の資料を参考とした。特に、分野（大分類）の設定等に当たっては、米国の産業界、大学関係者の参画により取りまとめられた報告書「The National Academies of Sciences : Data Science for Undergraduates Opportunities and Options」（2018年5月）を参考とした。

さらに、文部科学省、経済産業省、データサイエンス関係団体、放送大学及びデータサイエンス教育について先駆的な取組を有する公・私立大学との意見交換を行うなど、多方面からの意見収集に努めた。資料、意見収集について、第1次報告作成時のカリキュラム分科会の主査丸山祐造教授（現神戸大学）をはじめ各委員に感謝申し上げる。

- ・ データサイエンティスト協会：データサイエンススキルセット
- ・ 統計教育連携ネットワーク：コアカリキュラム
- ・ R. D. De Veaux et al.: Curriculum Guidelines for Undergraduate Programs in Data Science, Annual Review of Statistics and Its Application, pp.15-30 (2017)
- ・ American Statistical Association: ASA Statement on The Role of Statistics in Data Science (2015)
- ・ 日本学術会議：大学教育の分野別質保証のための教育課程編成上の参照基準：数理科学分野（2013）
- ・ 日本学術会議：大学教育の分野別質保証のための教育課程編成上の参照基準：統計学分野（2015）
- ・ 日本学術会議：大学教育の分野別質保証のための教育課程編成上の参照基準：情報学分野（2016）
- ・ 日本学術会議：ビッグデータ時代に対応する人材の育成（2014）
- ・ 情報・システム研究機構 ビッグデータの利活用に係る専門人材育成に向けた産学官懇談会：ビッグデータの利活用のための専門人材育成について（2015）

Ⅱ データサイエンス

1 データサイエンス教育

データサイエンスは科学や産業に革命を起こす新しいアプローチとして注目されている。今や多くの研究分野のみならず社会の至るところで大規模で網羅的なデータが急速に蓄積されるようになり、データやその分析結果を利用することが可能になりつつある。その結果、今後はデータこそが経済、社会及び日常生活の発展と新しい価値創造の源泉になると考えられている。データサイエンスはそのようなデータ中心主義に立脚した科学的方法論である。アメリカ国立科学財団（National Science Foundation, NSF）の StatSNSF² によればデータサイエンスは、「データの設計、取得、管理、解析及びデータからの推論を目指す科学」とされている。

産学官懇談会³の報告書で指摘されたように、データから新しい価値を生み出すことができる「棟梁レベル」をはじめとするデータサイエンスの専門人材を育成することの重要性は言うまでもない。同時に、専門人材が生み出す成果の受け手となる国民全体のデータリテラシーを醸成することが、我が国においてデータサイエンス、A I を活かして超スマート社会を実現していくために必要不可欠だと考えられる。したがって棟梁レベルからデータリテラシーレベルまで裾野の広い人材育成が望まれる。

特にデータリテラシーとしては、データサイエンスやA I の不適切な利用や、個人データの利用から生じうる倫理的問題を認識できるようになることが重要である。何故ならば、その認識が今後のデータ駆動型社会を安全・安心に過ごすために不可欠だからである。もちろんデータサイエンスの専門人材も、そのような側面を十分に認識した上で価値を創造することが強く期待される。不適切な利用の代表例としては、誤判断につながる不適切な可視化、相関と因果の混同が挙げられる。

【不適切な可視化】

データのグラフ（棒グラフ、折れ線グラフ、円グラフなど）により、データの重要な特徴を、比較的容易に他者に伝えることができる。そのために、グラフは学術論文や技術文書だけではなく、テレビを始めとするマスメディアで頻繁に登場する。ただし、発信元が予め想定した主張をサポートするために、しばしばバイアスを持って提示される。その代表例として、縦軸をゼロから始めずに長さを切り詰めた棒グラフが挙げられる。

【相関と因果の混同】

単なる X と Y の相関関係が、もっともらしいデータ分析結果に基づいて因果関係（原因 X が結果 Y に影響を与えるかのようにマスメディアで主張される事例は枚挙に暇がない。そのような主張に影響された個人が X を変えたとき、 Y に期待された効果が得られないだけでなく、副作用により健康に害を及ぼすかもしれない。また、法人の意思決定や政策形成の場で誤認識された因果関係が用いられれば、大きな利益損失や税金の無駄遣いを招くであろう。

² <https://www.nsf.gov/attachments/130849/public/Stodden-StatsNSF.pdf>

³ 2015年7月30日 情報・システム研究機構 ビッグデータの利活用に係る専門人材育成に向けた産学官懇談会 報告書「ビッグデータの利活用のための専門人材育成について」

また、倫理的問題としては、「プライバシー保護のために匿名化されたデータベースにおいても、外部データベースとの照合などの操作により、個人が再識別される可能性」や「個人が再識別されない場合でも、データ分析の結果の解釈によって、年齢・民族・性などで層別された特定のグループについて、秘密が曝露されたり、差別が引き起こされる可能性」などが挙げられる。人権が侵害されたり個人の尊厳が損なわれることのないデータサイエンスやAIの利用を目指した国際的な取り組みは始まったばかりである（2019年6月につくば市で開催された20カ国・地域（G20）貿易・デジタル経済相会合など）。我が国でも2019年3月に統合イノベーション戦略推進会議で「人間中心のAI社会原則」⁴が決定された。特に教育に関する項で、全ての人間がデータに関する倫理の基礎を学ぶことの重要性が指摘されている。

今後の社会はますますデータ駆動型となるために、データから価値を創造できるデータサイエンスの専門家への需要は高まる一方である。さらに上記のように、データサイエンティストとして主体的にデータから価値を創出する業務に携わらない人にとっても、データサイエンスのリテラシーが重要である。結局、全ての大学等の卒業生に期待される能力やスキルにデータサイエンスの素養が新たに加わったのである。

2 データサイエンスのサイクルと重要な分野

データサイエンスは課題解決に向けたデータに基づく知識発展のサイクルと捉えることが重要である。日本学術会議提言（2014）⁵やR. D. De Veaux et al.（2017）⁶で指摘されているように、データサイエンスにおいては

- 課題抽出と定式化
- データの取得・管理・加工
- 探索的データ解析
- データ解析と推論
- 結果の共有・伝達、課題解決に向けた提案

というプロセスがあり、課題解決と同時にデータからの知識発見や新たな課題発見が起るために、知識発展のサイクルが生じる（図1）。

⁴ 人間中心のAI社会原則（2019年3月統合イノベーション戦略推進会議決定、2019年6月G20貿易・デジタル経済相会議閣僚声明）で述べられている人間に期待される能力及び役割等は以下のとおり。

○ 人間に期待される能力及び役割

- ・ AIの長所・短所の理解、AIの情報リソースとなるデータ、アルゴリズム、又はその双方にはバイアスが含まれること及びそれらを望ましくない目的のために利用する者がいることの認識
- ・ AIやデータのバイアスの認識（統計的バイアス、社会の様態によって生じるバイアス、AI利用者の悪意によるバイアス）

○ 以下の原則に沿う教育環境が全ての人に提供される重要性

- ・ 誰でもAI、数理、DSの素養を身につけられる教育システム
- ・ 全ての人が文理の境界を超えて学ぶ必要
- ・ 以下のことを認識させるリテラシー教育

データにバイアスが含まれることや使い方によってバイアスを生じさせる可能性があること等のAI・データの特性があること

AI・データの持つ公平性・公正性、プライバシー保護に関わる課題があること

⁵ 提言 ビッグデータ時代に対応する人材の育成 <http://www.scj.go.jp/ja/info/kohyo/pdf/kohyo-22-t198-2.pdf>

⁶ R. D. De Veaux et al. (2017). Curriculum guidelines for undergraduate programs in data science. Annual Review of Statistics and Its Application, p.15-30

<https://www.annualreviews.org/doi/abs/10.1146/annurev-statistics-060116-053930>

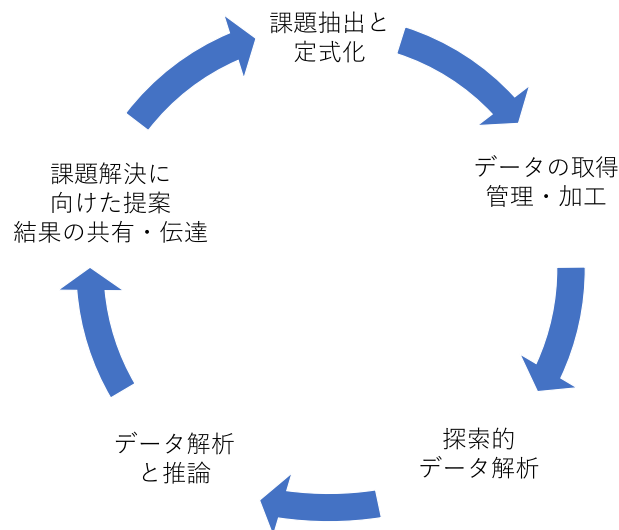


図1 データサイエンスのサイクル

そのようなデータサイエンスのサイクルを理解して実行する能力は、Google の Dr. Diane Lambert の言葉を借りれば「データと共に思考する能力」⁷と呼ばれる。また、アメリカの The National Academies of Sciences, Engineering, and Medicine によるデータサイエンスの学部教育に関する報告書「Data Science for Undergraduates Opportunities and Options」（以下、「DSU 報告書」という。）では「データ洞察力」と定義された能力に対応する。

DSU 報告書では、データ洞察力を学部教育において養うために、以下の 10 分野が重要としている。

- Ethical problem solving 倫理に配慮した課題解決
- Data description and visualization データの記述・可視化
- Data management and curation データの取得・管理・加工
- Statistical foundations 統計基礎
- Mathematical foundations 数学基礎
- Computational foundations 計算基礎
- Data modeling and assessment モデリングと評価
- Domain-specific considerations ドメイン知識の考慮
- Communication and teamwork コミュニケーションとチームワーク
- Workflow and reproducibility ワークフローと再現性

これらの 10 分野はサイクルの中で以下のように位置づけられる（図 2）。ただし、「倫理に配慮した課題解決」は全てのプロセスにおいて重要であり、また最後の「ワークフローと再現性」は、サイクルを持続させていくための作業手順の保持と再現性のことであり、各ステップには割り振られない。

⁷ Curriculum Guidelines for Undergraduate Programs in Statistical Science (2014)
<https://www.amstat.org/asa/files/pdfs/EDU-guidelines2014-11-15.pdf>

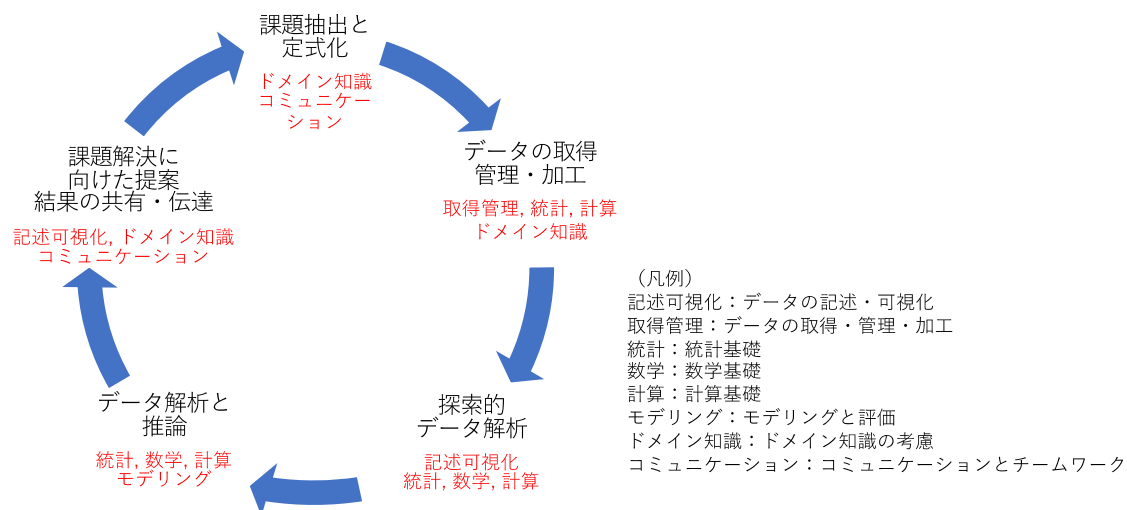


図2 データサイエンスのサイクル

3 カリキュラム分科会の取り組み

データサイエンス教育で先行する欧米では、600 を超えるデータサイエンス学部や教育プログラムがスタートしている。日本でも 2017 年度スタートの滋賀大学を皮切りに、データサイエンス学部を設置する大学がいくつか現れている。これらのデータサイエンス教育を推進する学部においては、データサイエンスのサイクル全般にわたって活躍するデータサイエンティストを養成する観点から、DSU 報告書で言及された 10 分野全てをカバーした教育が展開されることが重要であろう。一方で、データ駆動型社会のリテラシーとして、全国の全ての大学生・高専生もデータサイエンスの基礎を学ぶことが期待されている。時間的な制約などから、初級レベルの教育内容として全ての分野を考慮に入れるのは現実的ではない。しかしながら、データサイエンスのサイクルに主体的に関わるような職業に就かない場合でも、全ての国民がデータ駆動型社会の裾野となるとともに、データ分析の受け手としてデータサイエンスの成果を享受しつつ人生を生きていけるようになることが重要である。

そのような観点から、数理・データサイエンス教育強化拠点コンソーシアム カリキュラム分科会では、全ての大学生・高専生が身につけておくべきデータサイエンスのリテラシーを第 1 次報告として取りまとめ、2019 年 11 月に公表した。2020 年 4 月にコンソーシアムが公表した「数理・データサイエンス・AI（リテラシーレベル）モデルカリキュラム ～データ思考の涵養～」は、第 1 次報告を参照しつつ、産業界、私大等の取組状況を反映して取りまとめられたものである。

カリキュラム部会では、引き続き中級レベル、そしてさらに上級レベルへの橋渡し、それと同時に「AI 戦略 2019」に掲げられた「応用基礎レベル 25 万人／年」から「エキスパート 2,000 人／年」への橋渡しとなる、より高次のスキルセットの検討を進めてきた。検討に当たっては、第 1 次報告と同様に、特に以下の 7 分野に注目した。なお、DSU 報告書に示された「倫理に配慮した課題解決」については、法律・規制を含めて、「データの法規・倫理」としている。

- データの法規・倫理
- データの記述・可視化
- データの取得・管理・加工
- 統計基礎

- 数学基礎
- 計算基礎
- モデリングと評価

7 分野をそれぞれ大分類とみなし、その下に中分類・小分類と階層構造を設けている。
例えば「データの法規・倫理」では、

大分類 データの法規・倫理

中分類 情報倫理、情報セキュリティ

小分類 情報倫理・関連法規／情報セキュリティ

という構造である。

そして、小分類を単位として、身につけるべき事柄を**スキルセット**と**学修目標**として整理している。学修目標はスキルや概念のイメージが湧くように具体性を意識して文章で記述したものである。

なお、大分類等に対応させて科目設計を行うことは必ずしも意図していない。スキルセットと学修目標を参照し、各大学等の教育の特徴を活かしつつ、教育資源等も考慮しながら、各大学等において、科目編成やシラバス作成等を行うことを想定したものである。

【中分類・小分類のリスト】

初級レベルのスキルを含む中分類・小分類のリストは以下のとおりである。

大分類	中分類	小分類	レベル
データサイエンスを学ぶ意義	(データサイエンスを学ぶ意義, 重要性の理解) データ駆動型社会 データに基づく課題解決, 意思決定の支援 データサイエンスのサイクル データサイエンスの様々な事例		★
データの法規・倫理	情報倫理, 情報セキュリティ	情報倫理・関連法規	★
		情報セキュリティ	★
	データに関連する法律・規制	個人のデータに関連する法規	★
		著作権法	★
		統計法	★
		不正競争防止法	★★
		医療ビッグデータ法	★★
	データサイエンスの倫理	倫理に配慮したデータ収集	★
		倫理に配慮したデータ活用	★
		データの匿名化	★
		データサイエンスに関する様々なバイアス	★
		公表バイアス	★
データの記述・可視化	データの記述	種々のデータ	★, ★★
		基本統計量	★
		相関関係	★, ★★
	データの可視化	グラフの構成要素	★
		統計グラフ	★
		分布の統計グラフ	★, ★★
		ダッシュボード	★★
		グラフィカルモデル	★★★
データの取得・管理・加工	データ取得とオープンデータ	日本や世界のオープンデータ	★
		オープンデータの取得	★
	データ管理とデータ形式	表形式のデータ	★
		その他のデータ形式	★
		データベース	★, ★★
	データの前処理	データクレンジング	★
		外れ値, 異常値, 欠損値	★
		データ加工	★
統計基礎	確率と確率分布	場合の数, 順列・組み合わせ	★
		確率	★
		確率分布, 確率変数	★, ★★
		主要な確率分布	★, ★★
		確率変数の漸近的性質	★★
	データ収集法と確率構造	標本調査	★
		誤差の扱い	★
		ランダム化比較試験	★
	推測統計	統計的モデル	★
		標本分布	★, ★★
		点推定, 区間推定	★, ★★, ★★★
		仮説検定の概念	★
		汎用的な検定	★★
		種々の検定	★★
		多重比較	★★★
		ベイズ統計的推測	★★★
	ベイズ理論	事後確率	★
		事前分布・事後分布	★★
		ベイズ的仮説検定	★★
		分布の推定	★★★
	計算統計	ブートストラップ	★★★
		疑似乱数	★★★
		サンプリング	★★★
		モンテカルロ積分	★★
数学基礎	線形代数	ベクトル	★
		行列	★, ★★

		データ記述と線形代数	★★
		固有値と固有ベクトル	★★, ★★★
		n 次元ユークリッド空間	★★, ★★★
		数値計算と線形代数	★★, ★★★
	微積分	多項式関数	★
		初等関数	★, ★★
		1変数関数の微分法	★, ★★
		1変数関数の積分法	★, ★★
		2変数関数の微分法	★, ★★, ★★★
		2変数関数の積分法	★, ★★, ★★★
		数値積分	★★, ★★★
	数列	データと数列	★
		数列	★, ★★
	演習	線形代数演習	★, ★★
		微積分演習	★, ★★
計算基礎	情報, コンピュータの仕組み	数と表現	★
		有効数字, 計算の誤差	★
		データ量の単位	★
		デジタル化	★
		文字の表現	★
		集合と命題	★
		論理演算	★
	データ構造	配列, リスト	★
		データフレーム	★
		連想配列	★
		スタック, キュー	★★
		グラフ, 木構造	★★★
	アルゴリズムとプログラミング	アルゴリズム	★, ★★, ★★★
		プログラミング	★, ★★, ★★★
モデリングと評価	モデリングによる課題解決	モデル作成	★
	教師あり学習	回帰分析	★, ★★
		判別分析	★★, ★★★
		正則化法とモデル選択	★★
		非線形モデリング	★★★
		決定木	★★
		ベイズ統計・モデリング	★★★
		強化学習	★★★
	教師なし学習	クラスタリング	★
		主成分分析	★★
		多様体学習	★★★
		カーネル密度推定	★
	時系列データ	時系列モデリング	★★, ★★★
	生存時間解析	生存時間データ	★★
		生存関数の推定	★★, ★★★
		生存時間データの比較	★★★
	質的データ解析	質的データの解析	★★★
	テキストデータ解析	テキストデータの数値化	★★★
		テキスト分析	★★★
	モデルの評価	モデル評価指標	★, ★★
		訓練データとテストデータ	★
	因果推論	因果モデル	★★
		グラフィカルモデリング	★★, ★★★
	深層学習・ニューラルネットワーク	ニューラルネットワークの仕組み	★★, ★★★
		深層生成モデル	★★★

4 スキルセットと学修目標

本節では、スキルセットと学修目標の位置づけや構成等について説明する。

(1) スキルセット

スキルセットは、データサイエンスの習得すべきスキルや理解すべき概念を、小分類を単位として関連するキーワードやレベルと合わせて整理したものである。レベルについては、初級から上級のレベル別に以下の3段階で整理している。

- ★ 専門を問わず、全ての大学生・高専生が身につけるべき初級レベル
- ★★ 数理・データサイエンス教育を推進する大学等の一般教育又は専門基礎教育レベル
- ★★★ 数理・データサイエンス教育を推進する大学等の専門教育レベル

例えば、大分類「データの法規・倫理」中分類「情報倫理、情報セキュリティ」のスキルセットは以下のとおりであり、各小分類に複数のスキルや概念が対応し、関連するキーワードとレベルと合わせて整理されている。

小分類	概要	キーワード	レベル
情報倫理・関連法規	情報倫理と関連法規の基礎	関連する法律（知的財産，個人情報の保護，不正アクセス行為の禁止等）	★
	情報社会で生活する上でのマナー，モラル，倫理の意義	マナー，モラル，倫理	★
情報セキュリティ	情報セキュリティの3要素	情報セキュリティの3要素，機密性・完全性・可用性	★
	情報セキュリティを確保するための対策	パスワード，個人認証，ソフトウェアのセキュリティ更新プログラム	★

キーワードは概要を補足し分かりやすくするために挙げており、概要に含まれる内容をキーワードによって全て網羅しているわけではない。

(2) 学修目標

学修目標では、全大学生・高専生へのデータサイエンス教育の普及・展開に向けて、初級レベルから大学等の専門教育レベルまで、習得すべきスキルを文章で記述している。中分類について、当該中分類のデータサイエンス全体における意義や重要性を記述し、各小分類に対して1つ以上のスキルを提示している。例えば、大分類「データの法規・倫理」中分類「データに関連する法律・規制」に対する意義や重要性、当該中分類の小分類「個人のデータに関連する法規」に関するスキルは以下のとおりである。なお、各小分類にはレベル（★、★★、★★★）を表示しており、小分類内で異なるレベルを有する概要については各スキルの説明文にレベル（★、★★、★★★）を表示している。

大分類「データの法規・倫理」中分類「データに関連する法律・規制」に対する意義や重要性

（改正）個人情報保護法と統計法はデータに関する基本的な法規である。前者については、高度情報化社会において個人の情報・利益の保護と個人情報の有用性のバランスを取ることの重要性を学ぶ。後者については、公的統計が行政利用だけでなく、社会全体で利用される情報基盤として極めて重要であり、根拠に基づく（データに基づく）政策立案の基礎であることを学ぶ。

中分類「データに関連する法律・規制」小分類「個人のデータに関連する法規」に関するスキル

★個人のデータに関連する法規

1. 我が国において、個人情報保護法が策定された背景を理解している。
2. 近年の飛躍的な情報通信技術の進展に伴い生じた、保護すべき個人情報の範囲の精緻化や、個人情報の適切な利活用の手続き、国境を越えた個人情報の流通への対応を目的とした個人情報保護法の改正の意義を理解している。
3. 改正個人情報保護法における、個人情報取扱事業者が個人情報を第三者に提供する場合の本人同意の必要性の意義、及び他の根拠法令を含めた例外規定の意義を理解している。
4. 改正個人情報保護法の「個人情報」「個人識別符号」「要配慮個人情報」「匿名加工情報」の概念を理解している。
5. EU一般データ保護規則（GDPR）や「忘れられる権利」など国際的な動きを知っている。

学修目標では、習得すべきスキルのイメージが湧くように具体性を意識して記述した。そのために学修目標の1項目とスキルセットの1項目は必ずしも対応しない。

5 高大接続

データサイエンス教育を全国の全ての大学等に展開するにあたり、高校までの学習内容に大きなばらつきがあることが問題となる。そのためにスキルセットと学修目標には高校での学習内容（数学、情報）をできるだけ含むようにして、高大接続を図っている。

数学

モデルカリキュラムの当面のターゲットは、現行の高校学習指導要領の下で学んだ学生である。そのために現行の数学科目の単元との対応をつけたリストを提供する。ただし、数学の単元は有機的に繋がっており、非対応の単元がデータサイエンスにとって重要でないことを意味しないことに注意されたい。

情報

現行の学習指導要領において、普通教育における教科「情報」は「社会と情報」と「情報の科学」の選択式である。データサイエンスに関連する内容は「情報の科学」に多く含まれるが、高校では「社会と情報」の方が圧倒的に教えられている。また入試で「情報」科目の試験を行う大学等は少なく、これらの理由から現行の学習指導要領の教科「情報」と対応付けるのは適当ではない。

一方で、周知されているとおり次期の学習指導要領における「情報 I」には大きな注目が集まっている。全ての高校生の必修となり、大学入学共通テストでの出題が検討されている。さらにデータサイエンスのリテラシーにあたる内容を多く含んでいる。既に述べたようにモデルカリキュラムの当面のターゲットは、現行の高校学習指導要領の下で学んだ学生であるが、数年後の高校生が必修として学ぶ内容は現在の大学生・高専生にも不可欠な学習内容であろう。そのような観点から次期学習指導要領「情報 I」との対応をつける。

高校学習指導要領とスキルセット及び学修目標との対応の詳細は、附属資料1を参照されたい。

Ⅲ スキルセット

※中分類、小分類に付記した番号は、学修目標との対応を示す。

大分類	中分類	小分類	概要	キーワード	レベル
データサイエンスを学ぶ意義	(データサイエンスを学ぶ意義、重要性の理解) 1. データ駆動型社会 2. データに基づく課題解決、意思決定の支援 3. データサイエンスのサイクル 4. データサイエンスの様々な事例				★
データの法規・倫理	1. 情報倫理、情報セキュリティ	1.1 情報倫理・関連法規	情報倫理と関連法規の基礎	関連する法律（知的財産、個人情報の保護、不正アクセス行為の禁止等）	★
			情報社会で生活する上でのマナー、モラル、倫理の意義	マナー、モラル、倫理	★
		1.2 情報セキュリティ	情報セキュリティの3要素	情報セキュリティの3要素、機密性・完全性・可用性	★
			情報セキュリティを確保するための対策	パスワード、個人認証、ソフトウェアのセキュリティ更新プログラム	★
	2. データに関連する法律・規制	2.1 個人のデータに関連する法規	個人情報保護法の概要	個人情報保護法、改正個人情報保護法（第三者提供、個人情報の定義の拡充）	★
			個人のデータに関連する国際的な動き	EU 一般データ保護規則 (GDPR)、忘れられる権利	★
		2.2 著作権法	著作権法の概要	著作権法の基本事項と意義、柔軟な権利制限規定と情報解析	★
		2.3 統計法	統計法の概要	統計法の基本事項と意義、基幹統計調査	★
		2.4 不正競争防止法	不正競争防止法によるデータ不正利用対策	限定提供データ、目的外利用、契約の重要性	★★
		2.5 医療ビッグデータ法	医療ビッグデータ法の概要	医療ビッグデータ法の意義と重要性	★★
	3. データサイエンスの倫理	3.1 倫理に配慮したデータ収集	データ収集対象者からの同意の必要性、オプトアウトの仕組み	インフォームドコンセント、オプトアウト	★
		3.2 倫理に配慮したデータ利活用	データ利活用に関する事例検討	POS データ、交通カードデータ、GPS データ、人事データ、就職活動データ	★
		3.3 データの匿名化	匿名化と仮名化の違い、個人の再識別の可能性	匿名化、仮名化、再識別、秘密の曝露や差別の誘引	★
		3.4 データサイエンスに関する様々なバイアス	データサイエンスに関するバイアス	データバイアス、アノテーションバイアス、標本選択バイアス、帰納バイアス	★
		3.5 公表バイアス	公表バイアス、出版バイアス	公表バイアスの危険性、公表バイアスを減らす取り組み	★
データの記述・可視化	1. データの記述	1.1 種々のデータ	データの分類	質的変数と量的変数、名義尺度、順序尺度、間隔尺度、比例尺度	★
			時系列データ	時系列データ、季節調整、移動平均	★
			様々な特徴を持つデータ	パネルデータ、クロスセクションデータ、コホートデータ	★
			画像データ	画像データの扱い（数値化）、ノイズ除去、ディープラーニング	★★
			音声データ	PCM データ、周波数解析	★★
			テキストデータ	自然言語処理、形態素解析、頻度解析、正規表現	★★
		1.2 基本統計量	分布の中心とばらつきに関する基本統計量	最頻値、中央値、平均値、四分位範囲、分散と標準偏差、5 数要約	★
			分布における相対的な位置の概念	標準化、標準得点（z 得点）、偏差値、外れ値	★
		1.3 相関関係	2 つの量的変数の相関関係	散布図、共分散、相関係数	★
			多変数データ（3 つ以上の量的変数）の 2 変数間の相関関係	散布図行列、相関係数行列	★
			2 つ以上の質的変数の相関関係	分割表、クロス集計表、リスク比、リスク差、オッズ、オッズ比	★
			質的データを含む変数の相関関係の尺度	相関比、連関係数、順位相関係数	★★
			相関関係と因果関係の違い	相関と因果、交絡、偏相関係数	★
			説明変数が 1 つの単回帰分析	最小二乗法、回帰分析、説明変数、目的変数、予測値、残差、決定係数	★

	2. データの可視化	2.1 グラフの構成要素	グラフの構成要素	色の効果や特徴、点の色・大きさ・形状への配慮、線の太さと様々な破線	★
		2.2 統計グラフ	データ全体の傾向や特徴を読み取るためのグラフ	棒グラフ、折れ線グラフ、円グラフ、帯グラフ	★
			より多くの情報を含む統計グラフ	積み上げ棒グラフ、集合棒グラフ、レーダーチャート	★
			グラフの作り方、見方のリテラシー	チャートジャンク、不必要な視覚的要素	★
			Excel によるグラフ作成	グラフの挿入、ピボットグラフ	★
		2.3 分布の統計グラフ	ヒストグラム、箱ひげ図、散布図などのデータの分布を示すグラフ	ヒストグラム、箱ひげ図、散布図（2変数）	★
			より多くの情報を含む分布の統計グラフ	バイオリンプロット、ヒートマップ（3変数以上の散布図）、統計地図	★★
		2.4 ダッシュボード	テーマに沿った複数の要約統計やグラフ	統計局の統計ダッシュボード	★★
データの取得・管理・加工	1. データ取得とオープンデータ	1.1 日本や世界のオープンデータ	日本政府のオープンデータの取り組みとオープンデータの意義	二次利用可能なルール、機械判読への適性	★
			オープンデータの普及に向けた国際的な動き	オープンデータ憲章（2013年G8サミット）	★
		1.2 オープンデータの取得	統計ポータルサイトからのデータのダウンロード	e-Stat, DATA.GO.JP データカタログサイト, Open Knowledge International	★
			メタデータの検討	メタデータ、データ取得の経緯、変数の特徴	★
	2. データ管理とデータ形式	2.1 表形式のデータ	表形式のデータ	csv, tsv	★
		2.2 その他のデータ形式	データ記述言語、SNSのつながりや鉄道路線などのデータ	XML, JSON, 離散グラフ、キー・バリュ形式である隣接リスト	★
		2.3 データベース	データベースの意義と運用・管理	データベース管理システム(DBMS)	★
			データモデル	概念モデル、論理モデル	★
			リレーショナルデータベース	正規化、選択、射影、結合、SQL	★★
			トランザクションと障害時回復	トランザクション、ACID特性、障害時回復	★★
			ビッグデータに対応したデータベース	ビッグデータ、CAP定理、BASE特性、結果整合性、NoSQL	★
	3. データの前処理	3.1 データクレンジング	データクレンジング、正規化、標準化	文字列の表記の揺れの吸収（文字列、数字、日付、時刻）	★
		3.2 外れ値、異常値、欠損値	外れ値、異常値、欠損値	外れ値、異常値（箱ひげ図、標準化得点）、データ欠損の要因、欠損値の補完	★
		3.3 データ加工	データ加工	部分集合の抽出、行の並び替え、新しい列の追加	★
統計基礎	1. 確率と確率分布	1.1 場合の数、順列・組み合わせ	場合の数、順列・組み合わせ	場合の数、順列・組み合わせ、階乗	★
		1.2 確率	確率の概念、加法定理などの基本的な定理、ベイズの定理	加法定理、条件付き確率	★
		1.3 確率分布、確率変数	確率変数や確率分布の概念、離散型・連続型の確率分布	確率変数、確率関数、確率密度関数、累積分布関数	★
			確率変数の平均（期待値）と分散	平均、期待値、分散、標準偏差	★
			積率に関する不等式や高次の積率の概念	チェビシェフの不等式、尖度、歪度、積率母関数	★★
			複数の確率変数に関する諸概念	同時分布、周辺分布、条件付分布、共分散と相関、独立	★
			独立な確率変数の和の性質	独立な確率変数の和、期待値と分散、たたみ込みの密度関数	★★
		1.4 主要な確率分布	ベルヌーイ分布、二項分布	ベルヌーイ分布、二項分布、期待値と分散	★
			正規分布	正規分布、期待値と分散、線形和と正規性の保存	★

			その他の主要な1変量確率分布 (1)	ポアソン分布, 幾何分布, 指数分布, 一様分布	★
			その他の主要な1変量確率分布 (2)	対数正規分布, ガンマ分布, ベータ分布, 超幾何分布, 負の二項分布	★★
			2変量確率分布の視覚的理解	2変量正規分布, 確率密度関数の鳥瞰図	★
			多変量確率分布 (多変量正規分布, 多項分布)	多変量正規分布, 周辺分布と条件付き分布の正規性, 平均ベクトルと共分散行列	★★
		1.5 確率変数の漸近的性質	大数の弱法則と中心極限定理	大数の法則, 中心極限定理, 確率収束, 分布収束	★★
	2. データ収集法と確率構造	2.1 標本調査	標本調査と主要な標本抽出法	単純無作為抽出法, 層化抽出法, クラスター抽出法, 2段抽出法	★
			統計量を手がかりとした母集団の特性の推測	母数と統計量 (母平均と標本平均) の区別, 標本分布	★
		2.2 誤差の扱い	誤差の発生メカニズムの理解	系統誤差, 偶然誤差	★
		2.3 ランダム化比較試験	ランダム化比較試験, 根拠に基づく医療・政策立案	実験研究と観察研究, ランダム化比較試験, A/B テスト	★
	3. 推測統計	3.1 統計的モデル	統計的モデルと母数 (パラメータ) の概念	統計的モデル, 母数, パラメータ	★
		3.2 標本分布	独立同一分布 (i. i. d.) に従う確率変数の概念	独立同一分布, 標本平均, 標本分散	★
			i. i. d. 正規確率変数における標本平均・標本分散やそれらに派生する分布	カイ二乗分布, 標本平均と標本分散の独立性, t 分布, F 分布	★★
		3.3 点推定, 区間推定	点推定. 特にモーメント法, 最尤法による推定量の構成	モーメント法, 最尤法	★
			推定量の性質. 特にバイアス, 不偏性, 平均二乗誤差	バイアス, 不偏推定量, 平均二乗誤差, バイアス分散分解	★★
			推定量の評価	一貫性, 有効性, フィッシャー情報量 (クラメル-ラオの不等式)	★★★
			信頼区間の構成. 信頼係数の理解	信頼区間と信頼係数, 平均の区間推定, 比率の区間推定	★★
		3.4 仮説検定の概念	基本的な仮説検定 (二項分布の成功確率) を通じた仮説検定の考え方	帰無仮説, 対立仮説, 2 種の誤り	★
			有意水準, 検出力, p 値など検定に関する指標	有意水準, 検出力, p 値, 検定統計量	★
		3.5 汎用的な検定	パラメトリックモデルにおける尤度比検定	尤度比検定	★★
			様々なノンパラメトリック検定	ウィルコクソン検定, 順位和, 並べ替え検定	★★
		3.6 種々の検定	(複数の) 母集団分布の平均に関する検定	t 検定, 一元配置分散分析	★★
			複数の因子を持つ平均に関する検定	二元配置分散分析, 交互作用	★★
			適合度検定と分割表等の独立性の検定	適合度検定, 独立性の検定, カイ二乗分布	★★
		3.7 多重比較	検定の繰り返しによる第一種の過誤の確率の増加. その補正方法	ボンフェロニ補正, ホルム補正, FDR	★★★
		3.8 ベイズ統計的推測	事前分布・標本分布・事後分布, 共役事前分布	事前分布, 共役事前分布, 事後分布	★★★
	4. ベイズ理論	4.1 事後確率	事後確率の計算	ベイズの定理	★
		4.2 事前分布・事後分布	事前分布・標本分布・事後分布, 共役事前分布	事前分布, 共役事前分布, 事後分布	★★
		4.3 ベイズ的仮説検定	事前分布を考慮した仮説検定	仮説検定, 事前分布, 事後分布, ベイズファクター, ベイズアプローチ	★★
		4.4 分布の推定	数値シミュレーションによる分布の推定, MCMC 法	マルコフ連鎖, メトロポリス・ヘイスティングス法, ギブスサンプリング法	★★★
	5. 計算統計	5.1 ブートストラップ	ブートストラップ. より一般にリサンプリングに基づく手法	復元抽出, 経験分布, リサンプリング, ブートストラップ	★★★
		5.2 疑似乱数	決定論的なコンピュータに確率変数を発生させる疑似乱数. 一様乱数	疑似乱数, 合同法, メルセンヌ・ツイスター法	★★★
		5.3 サンプリング	目標分布からの (独立な) 疑似乱数生成. 逆変換法, 棄却法	逆変換法, 棄却法	★★★

		5.4 モンテカルロ積分	モンテカルロ積分	モンテカルロ積分, 期待値や確率密度の正規化定数	★★
数学基礎	1. 線形代数	1.1 ベクトル	ベクトル	n 個の数値の組としての n 次元ベクトル ($n=2, 3, \dots$)	★
			ベクトルの基本的な演算	ベクトルの和, スカラー倍, 線形結合	★
			内積	内積, ベクトルの直交性, ノルム, 単位ベクトル	★
		1.2 行列	行列 (基本的な演算と積)	行列の和とスカラー倍, 積の非可換性, 結合則, 分配則	★
			特別な行列とその性質	正方形列, 単位行列, 転置行列, 対称行列, 三角行列, 直交行列	★★
			正方形列の逆行列と求め方	正方形列の逆行列, 行列とその逆行列の積の可換性	★★
			行列の基本変形とランク	基本変形, ランク, 簡約な行列	★★
			行列を特徴付ける量とその性質	行列式, トレース, 正則性	★★
		1.3 データ記述と線形代数	分散, 共分散とベクトルの内積	all-ones ベクトル, 偏差ベクトル, 2つの偏差ベクトルの内積	★★
			回帰分析と射影操作	射影行列, 回帰分析における予測値ベクトルと残差ベクトル	★★
		1.4 固有値と固有ベクトル	対称行列の固有値と固有ベクトル	対称行列の固有値, 固有ベクトル	★★
			対称行列の直交行列による対角化	対称行列の対角化, スペクトル分解, 直交行列	★★
			二次形式と正定値性	二次形式と (半) 正定値行列	★★
			特異値分解の定義と構成	特異値分解	★★★★
		1.5 n 次元ユークリッド空間	n 次元ベクトルと n 次元空間上の点との対応	n 次元ベクトル, n 次元空間上の点の表現	★★
			ベクトルの線形独立性, 線形従属性	線形独立, 線形従属	★★
			線形空間, 線形部分空間の基底と次元, 複数のベクトルの張る線形部分空間	線形部分空間と基底・次元, 行列のランクとその列空間の次元	★★
			連立一次方程式の解の判定と解法	同次方程式, 係数行列, 拡大係数行列, 解空間, 解の一意性	★★
			正規直交基底とその構成法	正規直交基底, シュミットの直交化, QR 分解	★★★★
			射影操作	射影と直交成分, 直交補空間, 射影行列, ベキ等性	★★★★
			線形不等式	線形不等式論 (Farkas の補題, 線形計画の双対性)	★★★★
		1.6 数値計算と線形代数	連立一次方程式の数値計算における行列の性質	LU 分解, QR 分解	★★
			大規模連立一次方程式の解法, 固有値解析の解法	反復法など	★★★★
	2. 微積分	2.1 多項式関数	二次関数, 多項式と基本性質	二次関数や多項式のグラフ, 四則演算	★
			二項定理と二項係数	二項定理, パスカルの三角形	★
		2.2 初等関数	指数関数, 対数関数と基本性質	指数関数の形の把握, 指数法則, 対数関数の形の把握, 対数法則	★
			三角関数と基本性質	三角関数の形の把握, 三角関数の公式	★
			自然対数の底と指数関数	自然対数の底, ネイピア数	★
			三角関数の逆関数	$\arctan$ などの逆三角関数	★★
			関数の極限の定義と計算	関数の極限	★
		2.3 1 変数関数の微分法	導関数と関数の増減	平均変化率, 微分係数と接線, 導関数	★
			初等関数の導関数	初等関数の導関数	★
			微分法の基本的な関係	合成関数の微分法, 積の微分法	★
			導関数と関数の極値と最大・最小	関数の最大・最小, 極大・極小と導関数	★
			初等関数のテイラー展開	テイラー展開	★★
			方程式の数値的な解法	方程式の数値的解法, 二分法, ニュートン法	★★
		2.4 1 変数関数の積分法	微分の逆演算と原始関数	原始関数, 不定積分	★
			定積分の計算	定積分と微分積分学の基本定理	★
			積分の基本的な関係式	部分積分, 置換積分	★

			広義積分	広義積分	★★
			広義積分によって定義される関数	ガンマ関数, ベータ関数	★★
		2.5 2変数関数の微分法	偏微分の定義	偏微分, 偏導関数, 接平面	★
			2変数関数の最大最小問題	最大最小問題, 極値と偏導関数の関係, ヘッセ行列	★
			2変数関数のテイラー展開	テイラー展開	★★
			最急降下法と最適化手法	最急降下法, (最適化手法としての) ニュートン法, 準ニュートン法	★★
			多変数の非線形方程式の数値解法	(非線形方程式の解法としての) ニュートン法	★★
			条件付き極値問題の解法	ラグランジュ乗数法, 条件付き極値問題	★★★
			不等式による条件付き極値問題	Karush-Kuhn-Tucker (KKT) 条件	★★★
			凸関数条件	凸関数 (定義, ヘッセ行列の (半) 正定値性との関係, 最適性条件)	★★
		2.6 2変数関数の積分法	長方形領域での重積分	重積分, 重積分 (長方形領域), 累次積分	★
			一般の領域での重積分	一般の領域での重積分 (縦線領域, 横線領域の重積分)	★★
			重積分における置換積分	変数変換とヤコビアン	★★
			広義重積分	広義重積分, ガウス積分, 正規分布, 正規分布の極座標変換による求積	★★
			一般の多変数変数における重積分	多重積分, 極座標変換	★★★
		2.7 数値積分	基本的な数値積分の手法	台形則, シンプソン法	★★
			モンテカルロ法による定積分計算	モンテカルロ法, 乱数生成	★★★
	3. 数列	3.1 データと数列	データの標本平均や分散の表現	標本平均, 加重平均, 幾何平均, 調和平均, 標本分散, ベクトルの内積	★
			基本的な数列	等比数列, 等差数列	★
		3.2 数列	数列の収束の速さの違い	収束の速さ	★★
			数列および数列の和の極限	数列の極限, 数列の和の極限	★
	4. 演習	4.1 線形代数演習	ベクトルや行列を対象とした手計算演習	ベクトルや行列を対象とした手計算演習 (内積, 逆行列, 固有値)	★★
			ベクトルや行列を対象とした計算機演習	R や Python を用いた計算機演習	★
		4.2 微積分演習	微積分に関する手計算演習	解析的な計算が可能な問題の手計算による演習	★★
			微積分に関する計算機演習	Excel, R, Python などの既存の関数, ライブラリを用いた数値計算	★
				数式処理の演習 (Python の Sympy パッケージなど)	
計算基礎	1. 情報, コンピュータの仕組み	1.1 数と表現	2進数, 基数変換, 負の数と2の補数表現	2進数, 基数変換, 負の数と2の補数表現, 2進数の加減算, 表現可能な数値の範囲	★
		1.2 有効数字, 計算の誤差	浮動小数点演算と誤差. 丸め誤差	有効数字, 仮数部, 指数部	★
		1.3 データ量の単位	ビット, バイトなどのデータ量の単位. SI 接頭辞を使った表現	データ量の単位 (ビット, バイトなど), SI 接頭辞 (k, M, G, T, P, m, μ , n, p など)	★
		1.4 デジタル化	アナログとデジタルの特徴, デジタル化のプロセス, 画像データ	量子化, 標本化, 符号化, 画像データ (ラスタ, ベクタ)	★
		1.5 文字の表現	文字コード (文字の数値による表現)	ASCII, Unicode	★
		1.6 集合と命題	集合と命題, ベン図, 真理値表などの基本的な考え方	真偽値, 否定, 論理和, 論理積	★
		1.7 論理演算	論理演算の考え方と基本的な演算, 及び真理値表の利用方法	論理回路, 論理ゲート, AND, OR, NOT, NAND, XOR	★
	2. データ構造	2.1 配列, リスト	要素数固定の配列. 要素数可変のリスト	メモリ, ベクトル, 行列, アドレス	★
		2.2 データフレーム	異なる形のデータをまとめた (R や Python で使用できる) 変数	データフレーム	★

		2.3 連想配列	項目（キー）と値（バリュー）を組とするデータを格納するデータ構造	連想，辞書，ハッシュ	★
		2.4 スタック，キュー	スタック（後入れ先出しのデータ格納），キュー（先入れ先出しのデータ格納）	push, pop, enqueue, dequeue	★★
		2.5 グラフ，木構造	グラフ（ノードとエッジで表現）．木構造（階層性を有するグラフ）	ネットワーク，可視化，分類，二分木，ヒープ	★★★
	3. アルゴリズムとプログラミング	3.1 アルゴリズム	フローチャートの記号と処理手順の表現，疑似コード，アクティビティ図	端子，処理，判断，矢印	★
			アルゴリズムの基本構造（順次構造，分岐構造，繰返し構造）	代入，計算，順次構造，選択構造，繰返し構造	★
			基本アルゴリズム（合計，探索，併合，整列）	バブルソート，幅優先探索，深さ優先探索	★
			応用アルゴリズム（貪欲法，分割統治法，動的計画法，再帰的アルゴリズム）	メタ戦略，貪欲法，局所探索，分割統治法，動的計画法，再帰的アルゴリズム，フローネットワーク	★★
			計算量，計算量のオーダー（線形オーダー，べき乗オーダー），計算回数	ビッグO記法，入力データ量，計算時間，ステップ数，最大次数	★★
			最適化	（準）ニュートン法，最急降下法，座標降下法，勾配法	★★★
		3.2 プログラミング	Python や R などのインタプリタ型言語	ソースコード，解釈実行，対話的実行環境	★
			C/C++などのコンパイル型言語	ネイティブコード（機械語），翻訳，実行	★★
			構文（プログラムの書き方の規則）と意味論（構文要素の解釈）の区別	変数，代入，文，計算	★
			データ型，扱うデータを格納するためその種類に応じた変数の宣言	文字型，整数型，浮動小数点型	★
			演算子と式．各データ型に適用できる演算の種類とその表記	四則演算，論理演算，式	★
			繰返し処理や場合に応じた処理を記述する制御構造	for, while, if	★
			計算処理単位をまとめる関数．その定義と呼出し	関数，引数，返値，再帰呼出し	★
			ファイルや標準入出力などに対するデータの読み書き	標準入出力，ターミナル，ファイル，print, scanf	★
			再利用可能な定型処理とリファクタリング	計算の構造化，モジュール化，リファクタリング	★★
			オブジェクト指向	オブジェクト指向，オブジェクト，クラス，カプセル化，継承，多相性	★★★
			プログラムの誤り（バグ）の有無を検証するデバッグ，テスト	デバッグ，テスト	★
			応用プログラミング（ツールやライブラリを用いた課題解決）	表計算，テキスト処理，Excel，ライブラリ，API	★
			データ分析	実データを用いたデータ分析（各種手法による分析）	★
	モデリングと評価	1. モデリングによる課題解決	課題解決フレームワーク	PPDAC サイクル	★
			モデリング	モデルの作成と検証	★
				モデルの解釈と有効性	★
		2. 教師あり学習	重回帰や多項式回帰の概念，主要な統計量	平均二乗誤差，正規方程式，最小二乗推定量，残差平方和，決定係数	★
			統計的推測	回帰係数の有意性検定（t 検定，F 検定），自由度調整決定係数，AIC，Stepwise 法	★★
			回帰診断	回帰分析の諸仮定の妥当性，多重共線性等の推測の信頼性	★★
			変数選択・モデル選択	自由度調整決定係数，交差検証法，AIC	★★

	2. 教師なし学習		ロジスティック回帰	最尤法によるパラメータ推定, ニュートン・ラフソン法, 推定された係数の解釈	★★
				オッズ比, 対数オッズ, 最尤推定, 尤度比検定, ワルド検定	★★
		2.2 判別分析	1 次式, 2 次式に基づく判別	線形判別分析, 二次判別分析	★★
			サポート・ベクタ・マシン	最大マージン判別, ソフトマージンと正則化, 線形サポート・ベクタ判別, カーネル, 非線形な分離曲面	★★★★
		2.3 正則化法とモデル選択	正則化法, 過学習と汎化誤差, モデル選択	バイアスとバリエーションのトレードオフの概念, リッジ回帰, lasso, スパースモデリング, チューニングパラメータの選択, 変数・モデルの次元削減	★★
		2.4 非線形モデリング	非線形関数によるモデリング (回帰, 判別)	多項式, 交互作用, スプライン関数, 一般化加法的モデル	★★★★
		2.5 決定木	決定木 (回帰, 判別)	決定木 (回帰, 判別), アンサンブル学習 (バギング, ランダムフォレスト, 勾配ブースティング決定木)	★★
		2.6 ベイズ統計・モデリング	ベイズ統計の考え方	ベイズの公式, 単純ベイズ, ベイズ判別 (各カテゴリーの事後確率), 階層ベイズ	★★★★
		2.7 強化学習	強化学習, マルコフ決定過程による学習	TD 学習, Q 学習, 動的計画法	★★★★
	3. 教師なし学習	3.1 クラスタリング	非階層クラスタリング (k -平均法)	k -平均法の手続き, 結果の頑健性 (初期値, 距離)	★
			階層クラスタリング	データの類似度, 階層クラスタリングの手続き, 樹形図, 結果の頑健性 (類似度, 併合法)	★
		3.2 主成分分析	低次元への射影	視覚化, 特異値分解, 主成分分析, 寄与率	★★
		3.3 多様体学習	低次元への射影, データ特徴量の抽出	多次元尺度法, Isomap, 局所性保存射影 (局所線形埋め込み, ラプラス固有写像)	★★★★
		3.4 カーネル密度推定	カーネル密度推定	ヒストグラム, カーネル関数	★
	4. 時系列データ	4.1 時系列モデリング	時系列データの傾向	トレンド (移動平均), 周期性	★★
			時系列データの特徴	強定常性, 弱定常性, 自己相関, 相互相関, スペクトル	★★
			モデリング	自己回帰 (AR) モデル, 自己回帰移動平均 (ARMA) モデル, 分散不均一 (ARCH) モデル	★★
			状態空間モデル	カルマンフィルタ, 粒子フィルタ・逐次モンテカルロ法	★★
			隠れマルコフモデル	隠れマルコフモデル, 信号処理, 音声認識	★★★★
	5. 生存時間解析	5.1 生存時間データ	生存時間データの考え方	打ち切り, 生存関数, ハザード関数	★★
		5.2 生存関数の推定	ノンパラメトリック法	カプラン・マイヤー法	★★
			パラメトリック法	指数分布, ワイブル分布, 対数ロジスティック分布	★★★★
			セミパラメトリック法	Cox 比例ハザードモデル	★★★★
		5.3 生存時間データの比較	仮説検定	log-rank 検定	★★★★
	6. 質的データ解析	6.1 質的データの解析	数量化理論	数量化 I~V 類	★★★★
	7. テキストデータ解析	7.1 テキストデータの数値化	単語処理 (分割, 辞書活用など)	形態素解析, ステミング, ストップワード, 単語埋め込み, ベクトル空間モデル, レンマタイゼーション	★★★★
		7.2 テキスト分析	単語埋め込み, 単語類似度	共起ネットワーク, 対応分析, トピックモデル, 言語モデル, カナ漢字変換, 機械翻訳, 文書検索, 文書要約	★★★★
	8. モデルの評価	8.1 モデル評価指標	平均誤差, 誤判別率	平均絶対誤差, 平均二乗誤差, 誤判別率	★
			平均二乗予測誤差	Mallows の C_p 基準	★★
			カルバック・ライブラー情報量	AIC, 情報量規準	★★
			モデルの事後確率	BIC, ベイズ, 周辺尤度, ベイズ因子	★★

			二値分類問題における混同行列に基づく評価	正解率(accuracy), 適合率(precision), 再現率(recall), 混同行列, F 値	★
				ROC 曲線とその AUC	★★
		8.2 訓練データとテストデータ	訓練データとテストデータ	訓練データとテストデータ, バイアスとバリエーションの概念, 汎化誤差, 過学習, 適合不足	★
			交差検証法	交差検証法 (leave-one-out, k-Fold), ハイパーパラメータ	★
	9. 因果推論	9.1 因果モデル	因果効果の推定	反事実モデル, 平均処置効果 (ATE), 処置群に対する平均処置効果 (ATT), 対照群に対する平均処置効果 (ATU)	★★
			マッチング, 傾向スコア	マッチング, 傾向スコアマッチング, Horovitz-Thompson 型推定量	★★
		9.2 グラフィカルモデリング	無向グラフ	偏相関行列, 距離行列, 類似度行列, 隣接行列, 無向グラフ	★★
				マルコフ確率場, グラフィカル lasso, 確率伝搬法	★★★★
			因果探索, 有向グラフ	ベイジアンネットワーク, 確率伝搬法, 線形非正規モデル (LiNGAM)	★★★★
				構造方程式モデリング	★★★★
	10. 深層学習・ニューラルネットワーク	10.1 ニューラルネットワークの仕組み	単純パーセプトロン	入力層, 出力層, シグモイド関数, シナプス結合	★★
			多層パーセプトロン	隠れ層, 誤差逆伝搬法	★★
			深層ニューラルネットワーク	勾配消失, 活性化関数 (ReLU, シグモイド関数, 動径基底関数, ウェーブレット), ドロップアウト, (確率的) 勾配降下法とそのアルゴリズム, バッチ正規化, 計算グラフ	★★★★
			畳み込みニューラルネットワーク (CNN)	フィルタ, プーリング, 音声認識, 画像解析, RNN, LSTM, GRU	★★★★
		10.2 深層生成モデル	変分オートエンコーダー	特徴抽出, ノイズ除去, 低次元化, 変分オートエンコーダー	★★★★
			敵対的生成ネットワーク (Generative Adversarial Network: GAN)	識別モデル, 生成モデル, 教師なし学習	★★★★

IV 学修目標

- データサイエンスを学ぶ意義
- データの法規・倫理
- データの記述・可視化
- データの取得・管理・加工
- 統計基礎
- 数学基礎
- 計算基礎
- モデリングと評価

データサイエンスを学ぶ意義

インターネットの社会への広範囲な浸透、情報通信・計測技術の飛躍的发展、A I 技術の活用領域の拡大により、大規模で網羅的なデータが急速に蓄積され、データそのものやその分析結果を利用することが可能になっている。その結果、日常生活の中でも、販売記録、購買履歴、満足度やランキングなどのデータを元に意思決定が行われるようになっていく。また、そうした個人の意思決定とその結果そのものが企業にとっては重要な情報であり、次の意思決定に利用される情報が更新される。このような情報循環は、社会を发展させ個人の日常生活を快適にする。学術の世界においても、専門分野に関わらずデータサイエンス、A I 技術の活用が加速しており、分野横断研究の活性化や新領域の創出が進んでいる。スポーツの世界においても、多くの種目でデータの収集・分析を元にした戦略構築が進んでいる。

データ駆動型への転換が急速に進む今日において、数理・データサイエンス・A I に関するリテラシーは、A I 時代を担う全ての学生にとって必要不可欠な基礎的素養となった。データサイエンスの必要性を事例を通じて理解し、継続的な学習に繋げていくことが極めて重要である。一方でそのようなデータが個人情報と紐付いている場合には、倫理的な問題が生じることも少なくない。こうした問題にどう対応すべきかについてはデータサイエンス教育の最初の段階から意識しておく必要がある。

このように、データサイエンスへの導入として、産業界・学術界等における実際の事例を参照しながら、データサイエンスは社会に多大な便益をもたらすこと、その一方で倫理的問題などの注意すべき側面があること、またそれらを深く理解するために（これから学ぶことになる）基礎が必要であることを学生に認識させる。

1. データ駆動型社会

1. 大規模で詳細なデータが急速に蓄積されるようになっている事例を挙げられる。
2. データやその分析結果の活用によって、経済、社会及び日常生活の様々な側面が大きく変化していることを理解している。
3. 学生の興味に応じて、また専攻分野によって、学術におけるデータ駆動型の成果の例を挙げられる。

2. データに基づく課題解決、意思決定の支援

1. 自分の生活の中で、データに基づいて意思決定した例を挙げられる。
2. 政策立案やビジネスの現場における、データを用いた課題解決や意思決定の意義を理解している。

3. データサイエンスのサイクル

1. データサイエンスは、目的の達成のために、課題設定、データ取得から始まる課題解決に至るまでのプロセスであることを理解している。特に課題解決と同時にデータからの知識発見や課題発見が起こるために、プロセスはサイクルをなすことを理解している。
2. データサイエンスのプロセスの各ステップの重要性を理解している。
3. サイクルの全体を通じて、倫理的問題を検討することの重要性を理解している。
4. サイクルの全体を通じて、再現性を担保することの重要性を理解している。

5. 他者と協同して課題解決を行う場面で、データやデータ分析に基づいた（根拠に基づいた）コミュニケーションの重要性や有効性を認識できる。

4. データサイエンスの様々な事例

1. データサイエンスを用いた取り組みの成功例を挙げられる。
2. 個人のデータの匿名化を行っても、再び識別され深刻な倫理的な問題が引き起こされる可能性があることを知っている。
3. バイアスのあるサンプルが誤った予測や分類、結果の解釈に繋がった例を挙げられる。
4. 不適切なグラフ、可視化が誤解をもたらす可能性があることを理解している。
5. 相関関係と因果関係の違いを理解している。

データの法規・倫理

1. 情報倫理、情報セキュリティ

データサイエンスを学ぶ際に、数学や統計などの理論を身につけるだけでなく、コンピュータで実データを分析する能力やその際の注意点を習得することが重要である。また、ネットワークを介したグループでの共同作業により、データサイエンスの課題解決を導くことも想定される。ネットワークに繋がったコンピュータを扱うにあたり情報倫理の基礎を学び、情報セキュリティへの理解を深める。

1.1 ★情報倫理・関連法規

1. 知的財産に関する法律、個人情報の保護に関する法律、不正アクセス行為の禁止等に関する法律などを含めた法規が情報社会で果している役割を理解している。
2. 情報社会で生活する上でのマナー、モラル、倫理の意義や、法を遵守することの重要性を理解している。
3. 現代情報社会では技術革新のスピードが非常に早く、法規や制度の対応が間に合わないケースがあること、そのために個人のモラルや倫理に基づく正しい対応が期待されることを理解している。

1.2 ★情報セキュリティ

1. 情報セキュリティの3要素である機密性・完全性・可用性の重要性を理解している。
2. 情報セキュリティを確保するための対策として、推測されにくいパスワードや生体認証などの個人認証の必要性、ソフトウェアのセキュリティ更新プログラムを適用する必要性、その提供が終了したソフトウェアを使い続けることの危険性を理解している。

2. データに関連する法律・規制

(改正) 個人情報保護法と統計法はデータに関する基本的な法規である。前者については、高度情報化社会において個人の情報・利益の保護と個人情報の有用性のバランスをとることの重要性を学ぶ。後者については、公的統計が行政利用だけではなく、社会全体で利用される情報基盤として極めて重要であり、根拠に基づく（データに基づく）政策立案の基礎であることを学ぶ。

2.1 ★個人のデータに関連する法規

1. 我が国において、個人情報保護法が策定された背景を理解している。
2. 近年の飛躍的な情報通信技術の進展に伴い生じた、保護すべき個人情報の範囲の精緻化や、個人情報の適切な利活用の手続き、国境を越えた個人情報の流通への対応を目的とした個人情報保護法の改正の意義を理解している。
3. 改正個人情報保護法における、個人情報取扱事業者が個人情報を第三者に提供する場合の本人同意の必要性の意義、及び他の根拠法令を含めた例外規定の意義を理解している。
4. 改正個人情報保護法の「個人情報」「個人識別符号」「要配慮個人情報」「匿名加工情報」の概念を理解している。
5. EU 一般データ保護規則 (GDPR) や「忘れられる権利」などの国際的な動きを知っている。

2.2 ★著作権法

1. 著作権法が、知的財産権の1つである著作権の範囲と内容について定めていることを知っている。
2. 著作権法の目的が、著作物や実演、放送などに関し、これらの文化的所産の公正な利用に留意しつつ、著作者等の権利の保護を図ることで、文化の発展に寄与することであることを知っている。
3. デジタル化・ネットワーク化の進展に対応するために平成30年著作権法改正において規定の抽象度を高め適切な柔軟性を持たせた、柔軟な権利制限規定に関する基本的な考え方を理解している。

2.3 ★統計法

1. 統計法が、公的統計の作成及び提供に関する基本事項を定めていることを知っている。
2. 統計法の目的が、公的統計の体系的かつ効率的な整備及びその有用性の確保を図り、国民経済の健全な発展及び国民生活の向上に寄与することであることを知っている。
3. 公的統計が行政利用だけでなく、社会全体で利用される情報基盤として位置付けられることの意義を理解している。
4. 国勢調査などの基幹統計調査は、公的統計の中核となる基幹統計を作成するための特に重要な統計調査であり、正確な統計を作成する必要性が特に高いことを理解している。

2.4 ★★不正競争防止法

1. 不正競争防止法の目的が、事業者間の公正な競争の促進およびこれに関する国際約束の実施を確保することにより、国民経済の健全な発展に寄与することであることを知っている。
2. 平成30年の不正競争防止法改正において導入された限定提供データについての規定に関し、限定提供データとして法律の保護を受けるための3つの要件を理解している。

2.5 ★★医療ビッグデータ法

1. 医療ビッグデータ法（次世代医療基盤法）の目的が、医療分野の研究開発のための匿名加工医療情報に関し、健康・医療に関する先端的研究開発及び新産業創出を促進し、健康長寿社会の形成に資することであることを知っている。
2. 医療ビッグデータ法における医療情報および匿名加工医療情報の定義を理解している。
3. 医療ビッグデータ法では、オプトインのほか、丁寧なオプトアウトも認められていることを理解している。
4. 認定匿名加工医療情報作成事業者の認定の基準および匿名加工情報の作成等にあたっての義務を理解している。

3. データサイエンスの倫理

データサイエンスを応用し多種多様で大量のデータを分析した成果は、社会に多大なる便益をもたらす。その一方で、深刻な倫理的な問題を引き起こす可能性があることが指摘されている。例えば、個人情報情報の曝露や、プロファイリング等の自動処理による年齢・民

族・性などで層別された特定のグループへの差別や人権侵害である。ここでは特に、倫理に配慮したデータ収集や利活用、匿名化されたデータであっても個人情報曝露するリスクがあること、またデータサイエンスに関連する様々な偏り（バイアス）について具体例を通じて理解を深める。

3.1 ★倫理に配慮したデータ収集

1. データ収集の目的、データの活用方法がデータ収集対象者に説明されたうえで同意を得ていることが倫理的に望ましいことを理解している。
2. データ収集対象者が分析対象から外れることを希望したときにオプトアウトを実現できる仕組みの必要性を理解している。

3.2 ★倫理に配慮したデータ利活用

1. データの利活用において、調査客体が他者に知られたくない変数（調査項目）が国や文化によって異なること、より一般に倫理的基準が国や文化によって異なること、を理解している。
2. データ利活用に関する様々な事例における、成功・失敗の倫理面からの要因を理解している。

3.3 ★データの匿名化

1. 個人情報を含んだデータの匿名化と仮名化の違いを理解している。
2. 氏名などの個人情報の代わりに研究用 ID などを用いて仮名化されたとしても、外部データベースとの照合などの操作により、個人が再識別されうることを具体例をもとに理解している。
3. 個人が再識別されない場合でも、データ分析の結果の解釈によって、年齢・民族・性などで層別された特定のグループについて、秘密が曝露されたり、差別が引き起こされる場合があることを理解している。

3.4 ★データサイエンスに関する様々なバイアス

1. データバイアス、アノテーションバイアス
 - a. データサイエンスにおいて、人手でタグをつけたデータを正解として教師あり学習が適用される場合があることを知っている。その際にタグ付けをする人の認知バイアスなどにより、特定のグループに対する差別が生じうることを理解している。
 - b. 推薦システムにおいて、あるアイテムに関する評点をつける場合、画面に表示されているその時点での平均点などに影響される傾向が見られること、またその繰り返しにより将来の平均点などにバイアスが生じることを理解している。
2. 標本選択バイアス
 - a. 得られた標本が母集団から偏りを持って抽出されている場合には、たとえ適切なデータモデリング手法で分析されても、信頼できない結果しか得られないことを理解している。
 - b. 既に取得されたデータは過去の状態を反映していることを理解している。特に将来の母集団の特性値を予測したい場合に、過去のデータが未来の母集団からの偏りのないサンプルと見なせるかどうかの検討が重要であることを理解している。

- c. 大学入試の成績と入学後の学業成績の関係を調べる場合に、不合格で入学していない個人のデータは原理的に欠損していることなどを例に、標本のバイアスや分析の限界を理解している。
 - d. 幹部社員の男女比が極端に偏っているような企業における採用活動で、現社員に基づくデータ分析をもとに就職希望者の適性を判断するような例を題材として、過去のデータ分析結果を学習した場合は年齢・民族・性などで層別された特定のグループに対する差別が生じうることを理解している。
3. 帰納バイアス
- a. データサイエンスにおけるモデルは様々な仮定のもとに設計されること、それゆえに現実の近似に過ぎないことを理解している。
 - b. 解釈の容易さのために相対的に単純な（いくつかの変数を含まない）モデルが好まれること、また相対的に単純なモデルの方が汎化誤差（訓練データで推定したモデルをテストデータに適用したときの誤差）を小さくする傾向があることを知っている。
 - c. bで述べた特質のために、弱い効果しか持てないマイノリティの影響が無視されるなど、特定のグループに対する差別が生じうることを理解している。

3.5 ★公表バイアス

- 1. 解析結果の意義が小さいと判断されたときには、公表されないことも多いため、第三者が触れることができる解析結果は、研究者等によって意義が大きいと判断されたものに偏りがちであるという公表バイアス（出版バイアス）の存在を理解している。
- 2. 公表バイアスのために治療法や薬の有効性と安全性に関する誤認が起こり、多くの人の健康に悪影響を及ぼす危険性があることを理解している。
- 3. 解析データやロジックの公開、事前に調査内容を登録しておく制度（臨床試験登録制度）など公表バイアスを減らす国際的な取り組みが進んでいることを知っている。

データの記述・可視化

1. データの記述

データには量的変数や質的変数などそのままの形でデータ分析ができるものと、画像や音声などのマルチメディアデータのように、データ分析を行うために量子化（数値化）が必要なものがある。いずれのデータも適切な方法で要約して記述することで、それらのデータの分布の特徴を理解することができる。一変量データに関する標本平均や標準偏差などの基本統計量の概念、多変量データの相関関係を相関係数や回帰分析などを用いて記述する方法などを学ぶ。

1.1 種々のデータ

1. ★データの型として、質的変数、量的変数の違いを具体例を踏まえて理解している。さらにより細かい分類である名義尺度、順序尺度、間隔尺度、比例尺度の特徴を、具体例を踏まえて理解している。
2. ★時系列データの具体例を挙げられる。季節調整の基礎である移動平均の考え方を理解している。また GDP など多くの公的統計では、原系列から季節調整を行って得られる数値が公表されていることを知っている。
3. ★クロスセクションデータ、パネルデータ、コホートデータなどを通して、データを多面的に把握することの重要性を理解している。
4. ★★画像データを数値化（量子化）できる。さらに、ノイズ除去の方法やディープラーニングによる特徴量の抽出方法を理解している。
5. ★★音声データのデジタルデータ（PCM データ）化の意味を理解している。さらに、PCM データからスペクトルデータへの変換方法やスペクトル包絡による可視化の概要を理解している。
6. ★★テキストデータを集計する際の前処理（形態素解析）や、各種集計（単純集計、共起分析、コンセプト分析など）を理解している。

1.2 ★基本統計量

1. 平均値、中央値、標準偏差、四分位数、四分位範囲などの基本統計量の概念を理解している。
2. 標準得点（ z 得点）や偏差値など、分布における相対的な位置の概念を理解している。
3. 何らかの統計ソフトウェアでそれらの基本統計量を計算できる。

1.3 相関関係

1. ★2つの量的変数の線形関係の強さの指標である相関係数について、正負の符号や値の取りうる範囲などの特徴を理解している。
2. ★3つ以上の量的変数に関する散布図行列や、相関係数行列の各成分の意味を理解している。
3. ★単回帰分析について、係数の最小二乗法の意味を散布図に当てはめた回帰直線に基づいて理解している。
4. ★2つの質的変数（名義尺度、順序尺度）について、データセットから分割表を作ることができる。また、オッズ比などを踏まえて変数間の関係を理解している。

5. ★統計ソフトウェアを用いて、2つの質的変数から分割表（クロス集計表）を出力できる。また2つの量的変数の相関係数を計算できる。さらに単回帰分析を実行できる。
6. ★相関と因果の相違点、交絡の概念を身近な例を踏まえて理解している。
7. ★★質的データに対する相関の概念や各種指標（相関比、連関係数、順位相関係数）について理解している。

2. データの可視化

ビッグデータ時代において、データの可視化は大規模データから本質を取り出し、データ駆動型意思決定を行うための重要なスキルである。効果的な可視化により、言葉では伝えるのが難しいデータの重要な特徴やパターンを、他者に伝えることができる。またデータサイエンスのリテラシーとして、データの特徴を正確に可視化するための技術、及び他者の作った不必要な視覚的要素などを含むグラフに騙されないためのポイントを身につける。

2.1 ★グラフの構成要素

1. グラフにおける次のような色の効果、特徴を理解している。
 - a. 順序を持たない離散的な項目を区別するための質的な色の利用
 - b. 同系色の濃淡などによる、連続的な量的変数あるいは順序を持つ離散的な項目の表現
 - c. 特徴的な個体などを強調するための鮮明な色と不鮮明な色の対比
2. 色覚特性に配慮した配色でグラフを描ける。
3. 点の色、大きさ、形状に注意できる。（2カテゴリー以上に分かれる2変量データを同時に散布図に描くときなど）
4. 線の太さと様々な破線の使い分けができる。（複数の折れ線グラフを同時に描くときなど）

2.2 ★統計グラフ

1. 【棒グラフ】棒グラフは比例尺度の量を表すためのグラフであること、また棒の長さ（面積）が量に比例するように描くことを理解している。
2. 【棒グラフ】棒の長さ（面積）を量に比例させる原則のために、縦軸をゼロから始めずに長さを切り詰めることが誤解を生むことを理解している。
3. 【折れ線グラフ】特に時系列データについて、折れ線グラフが（横軸の）時間とともに数量が変わる様子を表現できることを理解している。
4. 【折れ線グラフ】気温などの具体例を想定しながら、縦軸を必ずしも0点から始める必要はないことを理解している。
5. 【円グラフ】円グラフは扇形の中心角の大きさに各カテゴリーの全体における割合を表すことを理解している。
6. 【帯グラフ】割合のカテゴリー間の比較には、円グラフよりも帯グラフの方が望ましいことを理解している。
7. 【積み上げ棒グラフ】積み上げ棒グラフは量と割合を同時に表現できることを理解している。
8. 【集合棒グラフ】各カテゴリーが有している複数の量的データを比較する際には、集合棒グラフが望ましいことを理解している。

9. 【レーダーチャート】レーダーチャートは複数の順序尺度のデータのバランスを表現できること、及びレーダーチャートの見方を理解している。
10. 【チャートジャンク】影、グラデーション、3D、矢印などの不必要な視覚的要素（チャートジャンク）が、誤解を招くことを知っている。
11. 【チャートジャンク】チャートジャンクを含むグラフに遭遇したとき、必要に応じて元データを参照するなどの判断ができる。
12. Excel による様々なグラフの作成方法、解釈の仕方を知っている。またピボットグラフにより様々な視点によるグラフを適宜作成できる。

2.3 分布の統計グラフ

1. ★【ヒストグラム】ヒストグラムに基づいて、「単峰か否か」「歪みがあるか」などのデータの分布の特徴を理解している。
2. ★【ヒストグラム】ビンの個数、階級の幅によって印象が変わることを理解している。
3. ★【箱ひげ図】代表的な分位点により分布を要約する箱ひげ図を用いて、データの分布の特徴を読み取ることができる。
4. ★【箱ひげ図】少数の要約統計量に基づいているために、例えば異なるヒストグラムの形となる2つのデータセットから、同じ箱ひげ図が作られる可能性があることを理解している。
5. ★【散布図】散布図をもとに、2つの変数が線形な関係を持つか否か、関係は強いのか否か等の基本的な検討ができる。
6. ★【散布図】2つの量的変数に対し、散布図と相関係数の大きさ・絶対値を対応させることができる。
7. ★【散布図】2カテゴリー以上に分かれる2変量データを同時に散布図に描くときに、点の色、大きさ、形状に注意できる。
8. ★【散布図】3つ以上の量的変数に関する散布図行列について、各成分の意味を理解している。また相関係数行列に基づくヒートマップから、強い線形関係を持つ2変数の組を抽出できる。
9. ★★【バイオリンプロット】バイオリンプロットは、箱ひげ図よりも詳細な分布を表現できることを理解している。
10. ★★【ヒートマップ】2次元ヒートマップから、2次元空間上で定義される量的データの分布を把握できる。
11. ★★【統計地図】地域ごとのデータの可視化には、統計グラフが適していることを理解している。

2.4 ★★ダッシュボード

1. 複数のグラフを加工、可視化するには、ダッシュボードが有用であることを知っている。
2. （総務省の）統計ダッシュボードを使いこなせる。

2.5 ★★グラフィカルモデル

1. データの関係性を適切に可視化したグラフ（有向グラフ、無向グラフ）を作成できる。
2. 複数の変数の相関関係、条件付き独立の関係を表す無向グラフを作成できる。
3. 因果推論に基づく有向グラフを作成できる。

データの取得・管理・加工

1. データ取得とオープンデータ

データの取得はデータサイエンスの第一歩である。オープンデータを巡る国内外の動きを理解して、関心のあるデータ、課題解決のために有用なデータを探し出してダウンロードできるようになる。同時にメタデータなどからデータ取得の経緯や変数の特徴を理解して、課題解決の目的に適したデータであるか検討できることも重要である。ただし、課題解決に活用できるオープンデータが常に入手可能ではないことに十分留意する必要がある。

1.1 ★日本や世界のオープンデータ

1. 日本政府のオープンデータの取り組みを知り、オープンデータの意義を理解している。特に、「二次利用可能なルール」「機械可読性」の重要性を理解している。
2. オープンデータを利用した国、地方公共団体、民間事業者の取り組みを一つ以上挙げられる。
3. 2013 年の G8 サミットで合意された「オープンデータ憲章」など、オープンデータの普及に向けた国際的な動きを知っている。

1.2 ★オープンデータの取得

1. e-Stat、DATA.GO.JP データカタログサイト（日本）、Open Knowledge International、国際連合（世界）などを通じて、関心のあるデータセット（csv 形式）をダウンロードできる。
2. ダウンロードしたデータのメタデータなどに基づいてデータの取得された経緯や変数の特徴などを理解している。

2. データ管理とデータ形式

データサイエンスでは、課題の発見からその解決に至るプロセスで、データが常に中心に据えられる。そのようなデータをファイルとして蓄積するためのデータの様々な形式を理解し、適切に管理する方法を学ぶ。

2.1 ★表形式のデータ

1. csv 形式が行と列からなる表形式のデータを表現できることを理解している。
2. e-Stat 等からダウンロードされたデータを、メタデータ部分を取り除くなどの前処理により、何らかのソフトウェアに読み込める csv ファイルに編集できる。

2.2 ★その他のデータ形式

1. Web API (Application Programming Interface) によるデータが提供される場合の標準的なデータ形式である、XML 形式（属性と要素の組み合わせ）や JSON 形式（JavaScript のデータ定義文をベースとした、簡易的なデータ定義言語）を知っている。
2. データ間の関係性やつながりを表す場合などは、表形式以外の方法でデータを表現した方が扱いやすいことを理解している。
3. SNS での人のつながりや鉄道の路線などのような関係性について、頂点と辺で構成された離散グラフとして表現できることや、離散グラフがキー・バリュー形式である隣接リストで表現できることを理解している。

2.3 データベース

1. ★【意義】データベースとは、ある目的のために収集した大量の情報を一定の規則に従って体系的に整理・蓄積し、検索や抽出に利用するための仕組みであることを理解している。
2. ★【運用・管理】システムの障害や不正アクセスから守るために、データベースを運用・管理するソフトウェアとしてデータベース管理システム（DBMS）が必要であることを理解している。
3. ★【データモデル】データベースのスキーマはデータベースの構造であることを理解している。
4. ★【データモデル】スキーマを決定していく作業をデータモデルの作成またはモデル化と呼び、概念設計、論理設計、物理設計の順に行われることを理解している。
5. ★★【リレーショナルデータベース】レコード（行）とフィールド（列）からなる表形式で表現されたテーブル（表）の集合としてデータを管理し、それらのテーブルを共通項目で関連づけられる（リレーションシップ）ことを理解している。
6. ★★【リレーショナルデータベース】テーブル（表）の分割をもとにデータの矛盾や重複などを無くす正規化により、効率的にデータ管理ができることを理解している。正規化されたテーブルでは、そのテーブルの中で一意に行を識別する主キーを持つことを理解している。
7. ★★【リレーショナルデータベース】テーブル（表）に対する3つの操作（関係演算）である「選択」「射影」「結合」を知っている。これらの操作は SQL（Structured Query Language）と呼ばれる言語を利用して行うことを理解している。
8. ★★【トランザクション】データベースへのデータの登録や更新の処理はトランザクションと呼ばれる単位で行われることを理解している。
9. ★★【トランザクション】トランザクション管理では、データに矛盾がなく整合性を保つため、データベースに書き込みがすべてうまくいった上で更新の処理を行うことを理解している。
10. ★★【ACID 特性】データベースのトランザクション処理を行う上で必要不可欠とされる4つの性質（Atomicity、Consistency、Isolation、Durability）を理解している。
11. ★★【障害時回復】データベースには障害から回復する機能が必要であること、データベースでは定期的にバックアップを取り、また処理の記録をログとして保存する必要性を理解している。
12. ★★【障害時回復】障害が発生した場合は障害の原因を取り除き、バックアップを取ったところの状態まで戻すことを理解している。
13. ★【ビッグデータに対応したデータベース】リレーショナルデータベース以外の代表的なデータベースとしてキー・バリュ型（キー・バリュ型）のデータベースがあり、特に大規模データの処理に活用されていることを知っている。また、他のデータ形式に対応したデータベースを含めて NoSQL と呼ばれること、NoSQL がビッグデータの活用に関連して注目されていることを知っている。
14. ★【ビッグデータに対応したデータベース】分散処理において満たすことができる要求が限られるという CAP 定理、またシステム全体に求められる BASE 特性の概要を理解している。

3. データの前処理

データ分析する前の前処理が、分析結果の精度や効率性に影響する。そのために適切な前処理がデータサイエンスにとって必要不可欠な工程であることを認識する。

3.1 ★データクレンジング

1. 名字や住所などを例として、文字列の表記揺れ対策の必要性を理解している。
2. 数字の表記の揺れを吸収する必要性を理解している。（整数部分と小数部分の区切りに「,」と「.」が使われることや、桁区切り文字の地域による違いなど）
3. 時系列データにおける日付、時刻の表記の揺れを吸収する必要性を理解している。

3.2 ★外れ値、異常値、欠損値

1. 箱ひげ図、標準化得点などをもとに、外れ値や異常値について検討できる。
2. 必要に応じて、データが取得された状況に立ち返って入力ミスなどを検討する必要性を理解している。
3. データ欠損が起こる要因をいくつか具体例をもとに理解している。（観測されるはずであったその値の大小が欠損に影響する場合など）
4. 標本平均などを用いた、欠損値の（自動）補完について検討できる。

3.3 ★データ加工

1. レコード（行）とフィールド（列）からなる表形式で表現されたデータ行列から、ある変数に注目し、その値に基づいてデータ行列の部分集合を抽出できる。
2. 行を並び替えられる（例えばある変数に注目して昇順にする）。
3. 変数の絞り込みによりデータ行列の部分集合を抽出できる。
4. 既存の列に四則演算を含む簡単な関数を適用して、新しい列を追加したデータ行列を作成できる。

統計基礎

1. 確率と確率分布

データを確率的誤差を持つものとして理解するという統計的推測の基本原則により、研究課題の考察、処方箋において不確実性の定量的な提示が可能となる。そのようなデータの背後に仮定される確率構造の基礎を学ぶ。

1.1 ★場合の数、順列・組み合わせ

1. 場合の数、順列・組み合わせの概念を理解している。またそれらに関する基本的な例題を計算できる。

1.2 ★確率

1. 様々な事象の起こりやすさを数値で表す確率の概念を理解している。また加法定理、条件付き確率、ベイズの定理に関する基本的な例題を計算できる。

1.3 確率分布、確率変数

1. ★確率変数の意味を理解している。
2. ★離散確率変数を特徴づける確率関数、連続確率変数を特徴づける確率密度関数の概念や性質を理解している。
3. ★確率変数の期待値（平均）や分散の定義を理解している。また簡単な例で期待値や分散を計算できる。
4. ★★チェビシェフの不等式を用いて、平均まわりのデータのばらつきを評価できる。
5. ★★確率変数の尖度や歪度の定義を理解している。また積率母関数からモーメントを計算できる。
6. ★2つの確率変数の共分散・相関係数・独立性の概念を理解している。
7. ★多変数の同時分布、周辺分布の概念を理解している。
8. ★★独立な確率変数の和の分布（またその期待値や分散）を求めることができる。

1.4 主要な確率分布

1. ★ベルヌーイ分布、二項分布の定義を理解している。またそれらの期待値と分散を知っている。
2. ★正規分布の様々なパラメータ（平均と分散）をもとに描かれた確率密度関数の複数のグラフをもとに、正規分布の平均の位置やばらつきの違いを理解している。
3. ★その他の重要な分布（ポアソン分布、指数分布、一様分布など）について、確率関数、確率密度関数のグラフに基づいて、その性質を理解している。またそれらの分布に従うと考えられる事象の具体例を挙げられる。
4. ★★応用上有用な分布（対数正規分布、ガンマ分布、ベータ分布、超幾何分布、負の二項分布など）について、活用場面を知っている。
5. ★2変量正規分布の確率密度関数の鳥瞰図に基づいて、2変数の相関の強さについて理解している。
6. ★★一般の多変量正規分布の周辺分布や条件付き分布も正規分布に従うことを知っている。

1.5 ★★確率変数の漸近的性質

1. 確率変数が無数に存在しているときの極限の概念（確率収束、分布収束）について理解している。
2. 独立かつ同一の分布に従う確率変数の標本平均が母平均に収束すること（大数の法則）を理解している。
3. 独立かつ同一の分布に従う確率変数の標本平均の分布が正規分布に近づくこと（中心極限定理）を理解している。

2. データ収集法と確率構造

データを収集するとき、調査の対象となる母集団をもれなく調べる全数調査は多くの場合不可能である。そこで、標本を抽出して調査し、その結果から母集団の性質を推測する標本調査が必要となる。そこに現れる確率構造といくつかの抽出方法を学ぶ。また、データ収集に関連して、「根拠に基づく医療」「根拠に基づく政策立案」のためのデータ収集デザインであるランダム化比較試験の考え方を学ぶ。

2.1 ★標本調査

1. 抽出された標本の調査結果から、母集団の状況をできるだけ正確に推測するという標本調査の目的を理解している。
2. 母集団の平均と標本の平均が必ずしも一致しないことをシミュレーションによって確認できる。
3. 単純無作為抽出法の原理と意義を理解している。
4. 母集団が大規模なケースでは、単純無作為抽出を行うことが困難であることを理解している。
5. 層化抽出法、クラスター抽出法、2段抽出法などの調査方法の特性を理解している。

2.2 ★誤差の扱い

1. 誤差は系統誤差と偶然誤差からなることを理解している。
2. 系統誤差は適切な処理により除去できることを理解している。

2.3 ★ランダム化比較試験

1. ある要因の効果を測定する場合に、ランダム化比較試験のデザインが注目している要因以外の影響を確率的に取り除く原理を理解している。
2. ランダム化比較試験では倫理的に許されないケースがあることを知っている。（喫煙の健康に与える影響を測定するために、ランダムに分けた2グループに対して、「10年間喫煙を強いる」「10年間禁煙を強いる」など）
3. インターネットマーケティング分野で普及しているA/Bテストは、ランダム化比較試験の考え方を基礎にしていることを知っている。

3. 推測統計

データサイエンスにおいては、しばしば統計的推測の立場からデータに基づく推論が行われる。統計的推測の考え方は、データを確率的誤差込みのものと捉えて、複数の確率分布を含む候補の集合の中からデータと整合的な分布を推測するというものである。その候補の集合はパラメータによって表現されており、統計的モデルと呼ばれる。データに基づいて統計的モデルのパラメータの値を推測する代表的な方法が、点推定、区間推定、仮説検定である。

3.1 ★統計的モデル

1. データの背後に複数の確率分布の候補を表現するという統計的モデルの考え方を理解している。

3.2 標本分布

1. ★統計的モデルのもとでは、標本平均、標本分散などの統計量が確率変数であり固定した値ではないことを理解している。
2. ★標本平均と標本分散が独立であることを理解している。
3. ★★統計的推測、仮説検定で重要となるカイ二乗分布、t 分布、F 分布の構成法について理解している。

3.3 点推定、区間推定

1. ★正規分布の母平均の推定において、標本平均や中央値など複数の推定量が考えられることを理解している。
2. ★複数の推定量が考えられる状況で、不偏性や平均二乗誤差といった推定量の良さに関する性質をもとに使うべき推定量を選択できる。
3. ★★平均二乗誤差はバリエーション（分散）とバイアスの二乗の和で表すことができることを知っている。
4. ★★★推定量の有用な性質である一致性と有効性を理解し、有効性を判断する重要な指標であるフィッシャー情報量を計算できる。
5. ★区間推定の概念及び信頼区間の意味を理解している。
6. ★★母平均に関する信頼区間、母比率に関する信頼区間を計算できる。

3.4 ★仮説検定の概念

1. 帰無仮説と対立仮説の概念を理解し、帰無仮説が棄却されること（対立仮説が正しいと判断すること）と帰無仮説が棄却されないこと（帰無仮説を棄却する十分な根拠がないこと）の意味を正しく理解している。
2. 二項分布において帰無仮説「成功確率が $1/2$ である」対立仮説「 $1/2$ でない」の場合の p 値を計算できる。また予め与えられた有意水準と比較して、帰無仮説を棄却するかどうかの判断ができる。
3. 正規分布の母平均の検定において、また片側検定・両側検定の両方で、正規分布の密度関数のグラフに基づいて p 値を理解している。また予め与えられた有意水準と比較して、帰無仮説を棄却するかどうかの判断ができる。

3.5 ★★汎用的な検定

1. 尤度比の意味を理解し、尤度比検定統計量を計算できる。
2. ノンパラメトリック検定の考え方（順位の等確率性など）を理解している。
3. 具体的なノンパラメトリック検定（ウィルコクソン検定、順位和検定、並び替え検定など）がどのような帰無仮説に対する検定かを理解し、実際に仮説検定を実施できる。

3.6 ★★種々の検定

1. 母集団の分布が正規分布に従う場合の平均に関する検定（1 または 2 標本における t 検定、多標本における一元配置分散分析）における必要な仮定を理解し、実際に仮説検定を実施できる。
2. 複数の因子をもつ実験について、各因子の水準の効果（交互作用効果を含む）に関する検定（二元配置分散分析）における必要な仮定を理解し、実際に仮説検定を実施できる。
3. クロス集計表における仮説検定（独立性の検定、適合度検定）における検定統計量を計算でき、実際に仮説検定を実施できる（帰無仮説の下での検定統計量の自由度を知っている）。

3.7 ★★★多重比較

1. 仮説検定を複数回実施した場合、仮説検定全体としての第 1 種の過誤確率は個々の仮説検定の第 1 種の過誤確率よりも高いことを知っている。
2. 仮説検定を複数回実施する際、仮説検定全体としての第 1 種の過誤確率を補正する方法（ボンフェロニの補正、ホルムの補正）を知っている。
3. 仮説検定を複数回実施する際、帰無仮説が棄却されたもののうち真の帰無仮説が棄却される割合（FDR）を調整する方法（BH 法、BY 法など）を知っている。

3.8 ★★★ベイズ的統計的推測

1. 観測されたデータにより事前分布を更新する（事後分布を得る）というベイズ的統計的推測の概念を理解している。
2. パラメータの事後分布の基づく点推定値（MAP 推定値、EAP 推定値）を計算できる。

4. ベイズ理論

観測されたデータに基づいて過去の事象や背景情報を推測することはデータサイエンスにおいて重要な課題である。過去の事象や背景情報の分布（事前分布）を与え、観測されたデータで条件づけることによりその分布を更新する（事後分布を得る）というアイデアをベイズ理論という。ベイズ理論を学ぶことにより、データから過去や未知の背景情報を推定することが可能となる。

4.1 ★事後確率

1. ベイズの定理を理解し、事後確率を計算できる。

4.2 ★★事前分布・事後分布

1. 事前分布と事後分布の概念を理解している。
2. 共役事前分布の概念、及び共役事前分布の具体例を知っている。
3. 観測されたデータに基づき、事後分布を計算できる。

4.3 ★★ベイズ的仮説検定

1. ベイズ理論を用いた仮説検定（パラメータの事前分布を仮定し、観測データから得られる事後分布に基づいて帰無仮説が棄却されるか否かを判断する方法）を理解している。
2. ベイズ的仮説検定において、帰無仮説を棄却する根拠となる指標（ベイズファクター）について知っている。

4.4 ★★★分布の推定

1. マルコフ連鎖の概念を知っている。
2. マルコフ連鎖モンテカルロ法（MCMC 法）によるパラメータ推定法を理解し、MCMC 法を用いて事後分布を推定できる。
3. MCMC 法におけるアルゴリズム（メトロポリス・ヘイスティング法、ギブスサンプリングなど）を知っている。
4. MCMC 法において重要なパラメータである burn-in、thinning や定常分布に達しているかどうかを表す指標（Gelman & Rubin の指標）を理解している。

5. 計算統計

計算機的发展により、複雑な事象をシミュレーションにより再現できるようになり、統計的推測においても数値シミュレーションは不可欠な技術となっている。乱数に基づく数値シミュレーションにより、未知の状態（分布）を再現することで、複雑な推定問題を解決することが可能となる。

5.1 ★★★ブートストラップ

1. ブートストラップリサンプリングの方法を理解している。
2. ブートストラップ法による統計的推測を理解している。
3. ブートストラップ法の様々な応用事例（バイアス補正、仮説検定、バギングなど）について知っている。

5.2 ★★★擬似乱数

1. 擬似乱数発生法（合同法、メルセンヌツイスターなど）について知っている。
2. 擬似乱数の性能の評価法を知っている。
3. メルセンヌツイスターの特徴を知っている。

5.3 ★★★サンプリング

1. 一様乱数から様々な分布に従う乱数を生成できることを知っている。
2. 逆乱数法によって乱数を生成できる。
3. 棄却法によって乱数を生成できる。

5.4 ★★★モンテカルロ法

1. モンテカルロ法により、各種シミュレーションを実施できる。
2. モンテカルロ積分により、期待値や正規化定数を計算できる。

数学基礎

1. 線形代数

多変数のデータに関する基本操作を学ぶ。データサイエンスでは、ベクトルによってデータを表現し、表形式のデータをデータ行列と見なして様々な線形演算を行う。それを理解するために、2次元ベクトルと2行2列型の行列の演算から始めて、様々なサイズのベクトルや行列の演算やその表現に慣れる。

1.1 ★ベクトル

1. 複数個 (n 個) の数値の組を n 次元ベクトルと見なせる。
2. 2次元ベクトルの和、スカラー倍、また2つの2次元ベクトルの線形結合を計算できる。
3. 2つの2次元ベクトルの内積を計算できる。
4. 3次元以上のベクトルについて、線形結合と内積を計算できる。

1.2 行列

1. ★行列 (基本的な演算と積)
 - a. 縦×横の表形式で並んだ複数個の数値を行列と見なせる。
 - b. 2×2 行列のスカラー倍と、2つの 2×2 行列の和を計算できる。
 - c. 2×2 行列どうしの積を計算できる。また、 2×2 行列と2次元列ベクトルの積、2次元行ベクトルと 2×2 行列の積を計算できる。
 - d. 3×3 行列や 2×3 行列など様々なサイズの行列のスカラー倍と和、及び様々なサイズの行列どうしの積を理解している。
 - e. n 人のクラスの p 科目のテストの点数からなる表形式の $n \times p$ データ行列について、各人の合計点の n 次元ベクトルや各科目の平均点の p 次元ベクトルが、行列とベクトルの積によって得られることを理解している。
2. ★★特別な行列とその性質
 - a. n 次の正方行列について、 n 次の正方行列のスカラー倍、和、積を計算できる。
 - b. 単位行列とは対角成分のみ1であり、他は0の正方行列である。 n 次の単位行列は、他の n 次正方行列と積をとっても同じ正方行列に等しいことが計算できる。
 - c. $n \times m$ 行列の転置行列が $m \times n$ 行列になることを理解し、実際に転置行列を作ることができる。
 - d. 三角行列を作ることができる。
 - e. 2×2 行列の対称行列、 2×2 行列の直交行列の具体例を実際に計算して確かめることができる。
3. ★★正方行列の逆行列と求め方
 - a. 2×2 行列の逆行列を求めることができる。
 - b. 行列とその逆行列は積を交換しても結果は等しいことを理解し、 2×2 行列について、これを確かめることができる。
 - c. 3×3 行列を与えられたときに、その逆行列を求めることができる。
4. ★★行列の基本変形とランク
 - a. 行列が与えられたとき、基本変形を計算できる。

- b. 基本変形を繰り返して簡約な行列に変形することができる。簡約な行列への変形結果は基本変形の順番にはよらないことが確かめられる。
- c. 行列を簡約な行列へ変形することで行列のランクを求めることができる。
- d. 3×3 行列を与えられたときに、その逆行列を求めることができる。
- 5. ★★行列を特徴付ける量とその性質
 - a. 2×2 行列及び 3×3 行列の行列式を求めることができる。また、具体的な n 次正方行列の行列式を求めることができる。
 - b. 正方行列についてトレースを求めることができる。
 - c. 正方行列を与えられたとき、その正則性を確かめることができる。

1.3 データ記述と線形代数

- 1. ★データと線形代数
 - a. サンプルサイズ n の一変量データを n 次元ベクトルと見なせる。また標本平均や加重平均を、データベクトルと加重係数ベクトルの内積と見なせる。
 - b. データベクトルが n 人の身長（単位 m）であるとき、単位を cm に変換したベクトルが元のベクトルのスカラー倍で表現できることを理解している。
- 2. ★★分散、共分散とベクトルの内積
 - a. データベクトルと平均から、偏差ベクトルを求めることができる。
 - b. 2つの偏差ベクトルの内積を用いて分散および共分散を計算することができる。
- 3. ★★回帰分析と射影操作
 - a. 射影行列を用いてベクトルの射影を求めることができる。
 - b. 回帰分析における予測値ベクトルと残差ベクトルに関する射影の関係を理解し、具体的に計算できる。

1.4 固有値と固有ベクトル

- 1. ★★対称行列の固有値と固有ベクトル
 - a. 2次の対称行列について、固有値と固有ベクトルを求めることができる。
 - b. 3次の対称行列について、固有値が全て異なるときに固有ベクトルを求めることができる。
 - c. 3次の対称行列について、固有値が重複している場合でも固有ベクトルを求めることができる。
 - d. 4次以上の対称行列について固有値を求めることができる。
- 2. ★★対称行列の直交行列による対角化
 - a. 2次の対称行列の異なる固有値に対応する固有ベクトルは直交することを利用して、直交行列によって対角化することができる。
 - b. 3次の対称行列を直交行列によって対角化することができる。
 - c. n 次の対称行列は直交行列によって対角化できることを理解している。
- 3. ★★二次形式と正定値性
 - a. 二次形式を対称行列を用いて表すことができる。
 - b. 対称行列の正定値性を確かめることができる。
- 4. ★★特異値分解の定義と構成
 - a. 行列の特異値分解を理解している。
 - b. ソフトウェアを用いて特異値分解を実際に計算できる。

1.5 n 次元ユークリッド空間

1. ★★ n 次元ベクトルと n 次元空間上の点との対応
 - a. n 次元ベクトルが n 次元空間上の点に対応することを理解している。
2. ★★ベクトルの線形独立性、線形従属性
 - a. 2次元ベクトルの線形独立性と線形従属性を確かめることができる。
 - b. 3次元以上のベクトルの線形独立性と線形従属性を確かめることができる。
3. ★★線形空間、線形部分空間の基底と次元、複数のベクトルの張る線形部分空間
 - a. 2次元ベクトルの全体を線形空間ととらえることができる。また、基底を作ることができる。
 - b. 線形空間を n 次元ベクトルの性質から理解している。
 - c. 与えられた n 次元ベクトルが生成する線形部分空間の基底を求めることができる。
 - d. 行列のランクとその列空間の次元の関係を理解している。
4. ★★連立一次方程式の解の判定と解法
 - a. 係数行列と拡大係数行列の基本変形によって連立一次方程式の解を求めることができる。
 - b. 連立一次方程式の解の全体が線形空間であることを理解している。
 - c. 解の一意性と係数行列の性質を理解している。
5. ★★正規直交基底とその構成法
 - a. 線形空間の基底が正規直交基底であることを確かめられる。
 - b. シュミットの直交化によって正規直交基底を構成することができる。
6. ★★射影操作
 - a. 射影と直交成分の関係を理解している。
 - b. 直交補空間を求めることができる。
 - c. 射影行列の性質、特にべき等性を理解している。
7. ★★線形不等式
 - a. Farkas の補題から線形計画の双対性を線形計画法に適用できる。

1.6 数値計算と線形代数

1. ★★連立一次方程式の数値計算における行列の性質
 - a. LU 分解によって連立一次方程式を実際に解くことができる。
 - b. 行列の QR 分解を計算することができる。
2. ★★大規模連立一次方程式の解法、固有値解析の解法
 - a. 大規模な連立一次方程式を解くには、共役勾配法などの反復法を用いることが有用であることを知っている。
 - b. 通常の QR 分解を用いて固有値を求める方法は、大規模な行列の場合には膨大な時間がかかることを理解しており、その際には計算量を抑えたアルゴリズムが必要となることを知っている。

2. 微積分

確率・統計・データモデリングに現れる関数についての主要な操作を学ぶ。1 変数関数は1変数の確率・統計の基本となり、2変数関数は多変数の確率・統計の基礎となる。微分はデータモデリングに使われる関数の局所的振る舞いを理解するための道具であり、積分は確率変数の期待値計算や区間領域の確率など大域的な性質を考察するための道具となる。

2.1 ★多項式関数

1. 二次関数について理解し、二次関数を用いて数量の関係や変化を表現できる。
2. 多項式関数の四則演算を計算できる。また与えられた多項式のグラフを描ける。
3. 二項係数と二項定理を理解している。

2.2 初等関数

1. ★指数関数、対数関数と基本性質
 - a. 指数関数（ネイピア数を底とする指数関数を含む）、対数関数の基本性質を理解している。またそれらのグラフを描ける。
 - b. 多項式と指数関数の、増加あるいは減少のオーダーの違いを理解している。
2. ★三角関数と基本性質
 - a. 三角関数の基本性質を理解している。またそれらのグラフを描ける。
 - b. 三角関数の加法定理を用いて三角関数を組み合わせた式を変形できる。
3. ★自然対数の底と指数関数
 - a. 自然対数の底（ネイピア数）の定義と実際の値を理解する。
4. ★★三角関数の逆関数
 - a. 三角関数 $\sin$, $\cos$, $\tan$ のグラフをもとにして、逆三角関数 $\arcsin$, $\arccos$, $\arctan$ のグラフを描ける。
5. ★関数の極限の定義と計算
 - a. 関数値の極限の概念を理解し、 $\lim$ の記法を身につけ、簡単な例で極限値を計算できる。

2.3 1 変数関数の微分法

1. ★導関数と関数の増減
 - a. 平均変化率によって関数の増加傾向、減少傾向を判断することができる。
 - b. 微分係数を平均変化率の極限値として求めることができる。また、微分係数を用いて関数のグラフの接線の式を決めることができる。
 - c. 導関数によって関数の増減を判断することができる。
2. ★初等関数の導関数
 - a. 初等関数の導関数を計算することができる。
3. ★微分法の基本的な関係
 - a. 合成関数の微分法を使って合成関数の導関数を計算できる。
 - b. 積の微分法を使って導関数を計算できる。
4. ★導関数と関数の極値と最大・最小
 - a. 導関数を用いて増減表を作ることができる。
 - b. 関数の極大値・極小値を求めることができる。
 - c. 関数の最大値・最小値を求めることができる。
5. ★★初等関数のテイラー展開
 - a. 初等関数のテイラー展開を与えられた次数まで計算することができる。
 - b. テイラー展開を用いて初等関数を n 次関数で近似することができる。
6. ★★方程式の数値的な解法
 - a. 二分法を用いて方程式の数値解を計算できることがわかる。
 - b. ニュートン法によって方程式の数値解を計算できることがわかる。また、注意すべき条件を理解している。

2.4 1 変数関数の積分法

1. ★微分の逆演算と原始関数
 - a. 微分の逆演算として定めた原始関数、不定積分を理解している。
 - b. 初等関数の原始関数、不定積分を計算できる。
2. ★定積分の計算
 - a. 微分積分学の基本定理を用いて初等関数の定積分を計算できる。
 - b. 定積分と面積の関係を理解し、関数のグラフで囲まれた図形の面積を定積分として計算できる。
3. ★積分の基本的な関係式
 - a. 部分積分を用いて複雑な不定積分、定積分を計算できる。
 - b. 置換積分を用いて複雑な不定積分、定積分を計算できる。
4. ★★広義積分
 - a. 広義積分の収束を理解し、広義積分を計算できる。
5. ★★広義積分によって定義される関数
 - a. 広義積分によって定義されるガンマ関数、ベータ関数の基本的な性質を理解している。

2.5 2 変数関数の微分法

1. ★偏微分の定義
 - a. 偏微分の定義を理解し、偏導関数を計算できる。
 - b. 2 変数関数のグラフが 3 次元空間内の曲面として表現できることを理解し、接平面の方程式を計算できる。
2. ★2 変数関数の最大最小問題
 - a. 2 変数関数の極値と偏導関数の関係を理解し、極値を求めることができる。
 - b. 2 変数関数の最大値と最小値を求めることができる。
 - c. ヘッセ行列によって極値判定ができる。
3. ★★2 変数関数のテイラー展開
 - a. 2 変数関数のテイラー展開を 2 次の項まで計算することができる。
 - b. 2 変数関数のテイラー展開を n 次の項まで計算することができる。
4. ★★最急降下法と最適化手法
 - a. 最急降下法によって 2 変数関数の最小値を探索することを理解している。
 - b. 2 変数関数の最小値を探索するために用いるニュートン法、準ニュートン法を理解できる。
5. ★★多変数の非線形方程式の数値解法
 - a. 非線形方程式の解法としてのニュートン法を理解できる。
6. ★★条件付き極値問題の解法
 - a. 条件付き極値問題の解法としてラグランジュ乗数法を用いることができる。
7. ★★不等式による条件付き極値問題
 - a. 不等式による条件付き極値問題の解法としての Karush-Kuhn-Tucker (KKT) 条件を理解している。
8. ★★凸関数条件
 - a. 凸関数の定義を理解し、凸関数であることを判定できる。
 - b. ヘッセ行列の正定値性との関係を理解している。
 - c. 最適性条件と凸関数の関係を理解している。

2.6. 2 変数関数の積分法

1. ★長方形領域での重積分
 - a. 長方形領域での重積分を累次積分として計算できる。
 - b. 累次積分の順番によって計算が容易になる場合があることを理解している。
2. ★★一般の領域での重積分
 - a. 一般の領域での重積分を累次積分として計算できる。
 - b. 累次積分の順番によって計算が容易になる場合があることを理解している。
3. ★★重積分における置換積分
 - a. 重積分における変数変換にはヤコビアンが現れることを理解し、計算できる。
4. ★★広義重積分
 - a. 広義重積分を理解している。
 - b. 正規分布の密度関数の実数上の積分について、極座標変換による求積ができる。
5. ★★ ★一般の多変数関数における重積分
 - a. 一般の多変数関数の多重積分を累次積分によって計算できる。
 - b. 多変数の極座標変換を理解している。

2.7. 数値積分

1. ★★基本的な数値積分の手法
 - a. 台形則を理解している。
 - b. シンプソン法を理解している。
2. ★★ ★モンテカルロ法による定積分計算
 - a. モンテカルロ法による定積分の近似計算法を理解している。
 - b. 乱数生成の手法を理解している。

3. 数列

行列の演算を成分毎に記述する場合や、一般にデータサイエンスの手法を記述する際に、数列の和の記号 Σ が頻出する。記法に慣れることで概念を深く理解することが期待される。

3.1 ★データと数列

1. 数列の和の記号 Σ の意味を理解している。
2. 「データの記述」の標本平均、標本分散、標本相関係数などを Σ を使って書ける。
3. 加重平均、幾何平均、調和平均を Σ を使って書ける。
4. ベクトルの内積を数列の和として計算できる。

3.2 数列

1. ★基本的な数列
 - a. 等差数列や等比数列などの簡単な数列について、一般項を表現したり第 n 項までの和を求めることができる。
2. ★★数列の収束の速さの違い
 - a. 数列の収束の速さの違いを理解している。
 - b. 等比数列と n 次関数による数列の収束の速さの違いを理解している。
3. ★数列および数列の和の極限
 - a. 具体的に与えられた数列の極限值を計算できる。

- b. 数列の和が求められるとき、その極限値を計算できる。

4. 演習

データサイエンスでは、行列演算、関数の最大化など線形代数、微積分に関する様々な計算が登場する。計算方法を理論的に理解するだけでなく、学生が手元の計算機で実際に計算できることが重要である。例えば Excel、R、Python の利用が想定される。同時に、理論を実際に応用する場面では数式処理を有効に活用する。

4.1 ★線形代数演習

1. 行列やベクトルを、配列として入力できる。
2. 既存の関数を用いて、行列やベクトルの和、差、積を計算できる。

4.2 ★微積分演習

1. 数値積分の関数の使い方を知っている。
2. 数値的最適化のための関数の使い方を知っている。

計算基礎

1. ★情報、コンピュータの仕組み

数値やデータをコンピュータで扱う場合の基礎として、2 進数による数の表現やデータ量の表し方、集合と論理演算の基本的な考え方などについて理解している。また情報のデジタル化を科学的に理解し、目的や状況に応じて適切にデータを取り扱う力を身につける。またコンピュータを用いて数値計算やデータ解析を行う際に、コンピュータ内部では 2 進数を大きな桁で非常に高速に論理演算することにより計算されること、また原理的に計算誤差が生じうること、などを知っていることは強みになる。

1.1 数と表現

1. 数を 2 進数で表現できる。2 進数の加減算ができる。

1.2 有効数字、計算の誤差

1. コンピュータでは定められたビット数のデータが扱われていることを理解している。
2. コンピュータでは表現できる値の範囲や精度が有限であることを理解している。
3. コンピュータは計算処理を有限の桁数で行うこと、またそのために計算処理には誤差が生じうることを理解している。

1.3 データ量の単位

1. ビット、バイトなどのデータ量の単位や、SI 接頭辞 (k, M, G, T, P, m, μ , n, p など) を使ったデータ量の表現を理解している。

1.4 デジタル化

1. アナログとデジタルの特徴を理解している。
2. コンピュータで扱われる画像データには、ラスタデータとベクタデータがあることを知っている。またそれぞれの特徴を理解している。
3. 標本化の精度や量子化のレベルによって、画像などのファイルサイズが違うことを理解している。

1.5 文字の表現

1. コンピュータの内部では、文字も記号も 2 進法のデータとして扱われ、文字コードと呼ばれる固有の番号が割り当てられていることを知っている。

1.6 集合と命題

1. 集合と命題に関する基本的な概念 (用語、記法、包含関係、和集合、積集合など) を理解している。
2. ベン図を用いて集合と命題に関する基本的な概念を理解している。

1.7 論理演算

1. 真理値表の基本的な考え方と利用方法を理解している。
2. 論理演算の考え方が分かり、真理値表を用いて基本的な演算を実行できる。

2. データ構造

コンピュータ内では、データを変数へ格納したり（代入）、変数から値を取り出す（参照）操作を繰り返しながら、計算処理を進めていく。特に複雑な計算プログラムを作成する場合に、簡易かつ効率的に変数を取り扱うために適切なデータ構造があることを理解している。

2.1 ★配列、リスト

1. 複数のデータを1つの変数名と添え字で管理する記述を理解している。
2. 探索アルゴリズムなどに応じて適切なデータ構造があることを理解している。

2.2 ★データフレーム

1. 異なる型のデータをまとめて管理する記述を理解している。
2. 特定の条件を満たすデータを抜き出す方法を理解している。

2.3 ★連想配列

1. 項目（キー）と値（バリュー）を組とするデータ構造を理解している。
2. 項目・値形式のデータが、データ間の関係を見出したり、データを検索する処理に有用であることを理解している。さらにデータベースに広く活用されていることを理解している。

2.4 ★★スタック、キュー

1. データ管理にスタックとキューというリスト構造があることを理解している。
2. スタックでは、後に入れたデータが先に取り出される（Last in First Out）ことを理解している。
3. キューでは、先に入れたデータが先に取り出される（First In First Out）ことを理解している。
4. 深さ優先探索はスタック、幅優先探索はキューのデータ構造が適していることを理解している。

2.5 ★★★グラフ、木構造

1. データ間の関係性を頂点（ノード）とその間をつなぐ辺（エッジ）にて表現できることを理解している。
2. 階層性を有しているグラフ構造が木構造であることを理解している。閉路のない連結グラフであるとも表現できる。
3. ネットワークやデータの可視化、分類や探索・ソートアルゴリズムに利用されることを理解している。

3. アルゴリズムとプログラミング

データサイエンスにおいては、アルゴリズム的な課題解決が不可欠である。明確に要件定義し、問題を分解し、アルゴリズム的な解法に至る効率的な戦略が必要となる。さらに適切な高水準言語によるプログラミングにより、解法を実装する必要がある。そのためにアルゴリズムとプログラミングの基礎的素養を身につける。またデータサイエンスで広く使われるExcel（VBA）、R、Pythonの基礎的スキルを身につける。

3.1 アルゴリズム

1. ★文章、フローチャート、アクティビティ図などによって表現されたアルゴリズムを、順次構造・選択構造・繰り返し構造といった基本構造を踏まえて理解している。
2. ★アルゴリズムによって処理の結果や効率に違いが出ることを、代表的なアルゴリズムの例を踏まえて理解している。
3. ★★応用アルゴリズムとして貪欲法、分割統治法、動的計画法などの問題を小さくして解く方法が計算機を利用する場合に有効であることを理解している。
4. ★★アルゴリズムがどれくらい高速に計算できるかを大雑把に表す指標である計算量を理解している。計算量は、入力データやモデル変数の数と計算時間の関係として表現することを理解している。
5. ★★★定めた目的関数を最大にするような変数を探索する最適化という考え方を理解している。最適値を効率的に探索する方法として、最急降下法、ニュートン法、座標降下法などが考案されていることを理解している。

3.2 プログラミング

1. ★アルゴリズムをプログラムとして表現する方法を理解している。例えばデータ型、演算子、変数、条件分岐、繰り返し、関数、ライブラリ、API を理解している。
2. ★作成したプログラムの動作確認や、不具合の修正の重要性を理解している。
3. ★Excel (VBA) 、R、Python のいずれかのプログラミング言語で単純なアルゴリズム（例えば数値の合計、平均、最大、最小）を表現できる。
4. ★プログラムを機械が理解できるようにあらかじめ翻訳した実行ファイルを作成するコンパイラ言語と、実行するタイミングでプログラムの 1 行ごとに翻訳しながら処理するインタプリタ言語があることを理解している。
5. ★★R のパッケージを用いて、様々なデータ分析を行える。
6. ★★Python のパッケージ・モジュールを用いて、様々なデータ処理、データ分析を行える。
7. ★★再利用可能な部分を関数化、モジュール化することや解読性が高まるようにソースコードを修正する作業がプログラム開発の工程で重要であることを理解している。
8. ★★★プログラムを各もの（オブジェクト）の組み合わせとして開発すること（オブジェクト指向）が重要であると理解している。関係する考え方として、クラス、継承、カプセル化などを理解している。

モデリングと評価

1. ★モデリングによる課題解決

様々な事象を適切にモデリング（データ構造を定式化）することにより、その構造・特徴の把握、予測、検証を行うことが可能となる。また、モデルを作成する流れ、作成したモデルの評価法を学ぶことで、より良いモデルを構築できる。モデリングによる課題解決における標準的な流れ・考え方、およびそれらの解釈に慣れ、モデリングに基づくデータからの知識獲得の重要性を学ぶ。

1.1 ★PPDAC（Problem-Plan-Data-Analysis-Conclusion）サイクル

1. PPDAC サイクルの流れを理解している。
2. PPDAC サイクルに従い、モデルの作成、評価、改善ができる。

1.2 ★モデリング

1. 作成したモデルについて、そのモデルを検証することの重要性を理解している。
2. 得られたモデルを解釈し、現実の事象を反映できているかを評価することの重要性を理解している。

2. 教師あり学習

教師あり学習（supervised learning）は、何らかの変数（目的変数）を別の変数（説明変数）で予測あるいは説明したい場合のモデリングである。代表的な手法は線形回帰およびロジスティック回帰であり、それぞれ目的変数が連続、2 値（つまり 2 つのカテゴリへの分類に対応）の場合に用いられる。これらに習熟することで、非線形回帰や多項（カテゴリカル）回帰といった、より複雑なモデリング手法を理解するための基礎を養う。

2.1 回帰分析

1. ★ 2 つの説明変数に対する重回帰分析における最小二乗法、回帰係数の最小二乗推定値、回帰式の当てはまりの良さの指標である決定係数などを、回帰平面を含む鳥瞰図に基づいて理解している。
2. ★自分で収集したデータあるいは e-Stat などダウンロードしたデータを、統計ソフトウェアに読み込み、回帰分析を実行して結果の解釈を行うことができる。特に t 値や p 値に基づいて回帰係数の有意性を検討できる。
3. ★散布図において説明変数 x と目的変数 y に非線形な関係が見られる場合に、説明変数の多項式関数による重回帰分析（例：切片、 x 、 x^2 による回帰分析）を実行できる。
4. ★★目的変数が離散値（特に 2 値）の場合には線形回帰が不適切であることを、散布図を通じて理解している。またロジスティック回帰モデルの意義を理解している。
5. ★★ロジスティック回帰モデルにおける回帰係数の標準的な推定法として最尤法を説明できる。
6. ★★ロジスティック回帰モデルで推定された回帰係数の解釈（説明変数が 1 単位増加した場合の目的変数への影響）を回帰分析の場合との違いを比較しながら理解している。

7. ★★自分で収集したデータあるいは e-Stat などダウンロードしたデータをソフトウェアに読み込み、ロジスティック回帰分析を実行して結果の解釈を行うことができる。特に Z 値や p 値に基づいて回帰係数の有意性を検討できる。

2.2 判別分析

1. ★★訓練データによる判別器の構成によって将来データのグループ分けができる。また、線形判別と二次判別を使うことが適切な状況をそれぞれ理解し、使い分けることができる。
2. ★★クラスタリングによるデータ分類法との違いを理解している。
3. ★★サポート・ベクタ・マシンにおけるマージン最大化（ハードマージン、ソフトマージン）、およびカーネルトリックの考え方を理解している。さらに線形判別、二次判別、サポート・ベクタ・マシン等により判別分析を実行できる。

2.3 ★★正則化法とモデル選択

1. データから良いモデルを得る方法としてモデル選択と正則化法があることを理解している。特に、過学習の回避法としての正則化法の考え方を説明できる。
2. モデルの適合度とモデル複雑度の折中を行っていることを理解している。
3. 具体的な回帰モデルで変数選択やリッジ回帰、LASSO 回帰を説明・実行できる。

2.4 ★★非線形モデリング

1. 線形回帰と非線形回帰の差異を理解し、データ解析における問題点とあわせて説明できる。
2. 多項式回帰、交互作用モデル、スプライン回帰、一般化加法モデルなどの典型的な非線形回帰モデルを挙げ、その推定方法を説明できる。

2.5 ★★決定木

1. 決定木の考え方・解釈を理解し、簡単な図示を通じてそれを説明できる。
2. 情報利得に基づいて入力変数の空間をなるべく簡単に分割することで出力を区別するという、決定木の特徴（決定木の解釈のしやすさ）を理解している。

2.6 ★★ベイズ統計・モデリング

1. 事前事象とデータから事後分布を求めて知識を発展させるためのベイズの公式を理解・説明できる。
2. パラメータを定数扱いする頻度論統計との違いを理解・説明できる。
3. ベイズ統計ではパラメータの事後分布やその点推定値（MAP、EAP）を導出・計算することが一つの大きな目的であり、そこから様々な結論が誘導されることを理解している。

2.7 ★★強化学習

1. マルコフ過程の遷移確率の概念を理解している。
2. 報酬最大化を逐次的に行うという、強化学習における基本的なアルゴリズム・更新公式を理解している。
3. TD 学習やモンテカルロ法といった様々な学習方式があることを知っており、具体的なものをいくつか説明できる。

3. 教師なし学習

教師なし学習 (unsupervised learning) は、目的変数のみから成るデータセットの学習法の総称である。クラスタリングは共通の特徴を持ったいくつかの集団にデータを分割するための手法であり、それには類似したデータを順次集団化にしてゆく階層的方法と、集団の数を決めてからデータを所属させてゆく非階層的方法がある。主成分分析は多次元データのばらつきに着目して新たな座標系を構築する手法である。加えて、座標系変換を通じた次元圧縮による可視化や、モデリングにより母集団構造を把握する手法なども存在する。これらの手法を用いて、データ構造を理解することの重要性を学ぶ。分布形を特定せずに推定を実行するノンパラメトリック法の基礎も学ぶ。

3.1 ★クラスタリング

1. k -平均法による非階層クラスタリングの反復手続きを理解し、それが有効な実例を挙げられる。
2. 自分で収集したデータあるいは e-Stat などダウンロードしたデータを、R や Python などのソフトウェアに読み込み、 k -平均法を実行して結果の解釈を行うことができる。
3. 初期値や距離を変えてみて、 k -平均法の結果の頑健性について考察できる。
4. 階層クラスタリングの手続きはデータ間の距離（類似度）とクラスターの逐次的な併合の繰り返しから成ることを理解し、説明できる。
5. 階層クラスタリングが有効な実例を挙げられる。
6. 樹形図（デンドログラム）の読み取り方と意義を理解している。
7. 自分で収集したデータあるいは e-Stat などダウンロードしたデータを、R や Python などの統計ソフトウェアに読み込み、階層クラスタリングを実行して結果の解釈を行うことができる。
8. データ間の距離（類似度）やクラスターの併合法の選択によって、結果が変わり得ることを確認できる。
9. 混合正規分布モデルに基づくクラスタリングと、EM アルゴリズムを用いたその推定方法を理解している。

3.2 ★★主成分分析

1. 標本共分散行列（実対称行列）の直交行列による対角化にもとづいて構成される主成分および主成分スコアの意味を理解している。
2. 累積寄与率を解釈できる。
3. 高次元説明変数の次元圧縮という観点から主成分分析を説明できる。
4. データ行列の特異値分解と、標本共分散行列の関係を説明できる。

3.3 ★★★多様体学習

1. 高次元複雑データ空間の座標系を変換して低次元ユークリッド空間へ落としこむという操作が有益な具体例を挙げられる。
2. 多次元尺度法や ISOMAP など、代表的な手法をいくつか理解し説明できる。

3.4 ★★カーネル密度推定

1. バイアス・バリエンストレードオフを数式で説明できる。
2. 微調整係数としてのバンド幅の役割とそのサンプルサイズとの関係を理解できる。
3. カーネル関数の例をいくつか挙げられる。

4. カーネル密度推定による母集団構造の把握や異常検知の手順を理解できる。

4. 時系列データ

現実のデータの多くは時間と共に変化する現象を観測した時系列であることを理解し、その特徴の表現の仕方、モデリングの方法、予測の方法を学ぶ。

4.1 時系列モデリング

1. ★★トレンド、周期・季節性や分散構造など、モデルの基本的な特徴量を理解している。
2. ★★AR, ARMA, ARMAX, GARCH など標準的なモデルとその推定方法を理解している。
3. ★★状態空間モデルの用途を説明でき、カルマンフィルタや逐次モンテカルロ法など関連する代表的な推定法・計算法を理解している。
4. ★★ソフトウェアを用いて時系列モデルの推定・予測を実行できる。
5. ★★隠れマルコフモデルの構造とその推定アルゴリズムを説明できる。

5. 生存時間解析

病気が治るまでの時間、機械が故障するまでの時間など、ある事象が発生するまでの時間に関する研究は重要な問題である。このような、ある事象が発生するまでの時間に関する分析法を生存時間解析という。個別の事象に対して、その発生時間を推定することは、様々な課題の解決につながる。

5.1 ★★生存時間データ

1. 数種類ある打ち切りデータの形式およびそれらの違いを理解している。
2. 生存関数とハザード関数による生存時間モデルを構成できる。

5.2 ★★生存関数の推定

1. パラメトリック（最尤法）、ノンパラメトリック（カプラン・マイヤー推定量）、およびセミパラメトリック（コックスの比例ハザードモデル）な推定法を理解し、違いを説明できる。また、打ち切りの有無による違いを説明できる。

5.3 ★★生存時間データの比較

1. ロジランク検定や一般化ウィルコクソン検定といった基本的な検定方式を理解している。

6. ★★質的データ解析

質的データはアンケートデータや属性データ、装置の ON/OFF など多くの場面で見受けられるが、質的データに対しては、連続的なデータに対する解析法をそのまま適用できないことが多い。しかし、若干の工夫を加えることで、質的データから連続的なデータに対する解析と同様の知見を得ることが可能となる。

1. 連続変量の場合と異なり、数値として観測できない質的（カテゴリーカル）データには目的に準じて特別な取り扱いが必要であることを理解している。
2. 基本的な数量化法（I 類～IV 類）を理解している。

3. 目的変数が質的データの場合の一般化線形モデルを説明できる（ロジスティック・多項回帰、ポアソン回帰など）。

7. ★★★テキストデータ解析

アンケートの自由記述欄のコメントなどは、これまで施策決定者が読み、その概要を主観的に判断し、施策の決定を行っていた。しかし、膨大なテキストデータを客観的に分析することで、読むだけでは把握しきれなかった様々な特徴を読み取ることが可能となった。今日の SNS の普及により膨大なテキストデータを容易に収集でき、それらの解析を通じて世の中の様々な傾向を把握できる。

7.1 テキストデータの数値化

様々なデータ前処理、またそれらに対応した辞書やリストを作成する必要があることを理解している。

7.2 テキスト分析

1. 形態素・構文の解析に引き続いて意味・文脈を解析するといった、基本的な流れを理解し説明できる。
2. テキストデータを定量化することで、文章のグルーピングや、注目されている項目の特定ができることを理解している。

8. モデルの評価

複数のモデルが考慮される場合、個々のモデルの性能を定量的に評価し、その値に基づいたモデル間での相対比較によって「良い」モデルが選択される。性能の評価指標は一通りではなく複数あることを理解し、回帰問題における平均二乗誤差、二値分類問題における混同行列といった概念を学ぶ。訓練データで学習させたモデルの性能を別のテストデータで試すという、教師あり学習における性能評価の基本的な考え方に慣れる。

8.1 モデル評価指標

1. ★回帰問題において、平均二乗誤差が一つの評価指標となることを理解している。
2. ★二値分類問題において、混同行列（の真陽性、偽陽性、真陰性、偽陰性）に基づく正解率（accuracy）、適合率（precision）、再現率（recall）などの評価指標を理解・説明できる。
3. ★★分布の離れ具合を測る指標としてカルバック-ライブラー情報量（ダイバージェンス）があり、それを使うことによって統計的モデルの評価基準 AIC が得られることを理解している。
4. ★★その他 Mallows の C_p や AIC, BIC など、基本的なモデル評価基準の考え方を説明できる。

8.2 ★訓練データとテストデータ

1. 訓練データをもとに特定のモデルのパラメータを推定し、その性能をテストデータにおける評価関数の値によって測るという考え方を説明できる。
2. 訓練データにおいて、当てはまりが良い複雑なモデルがテストデータにおいて必ずしも性能が良いとは限らない過学習（overfitting）を理解している。

3. 訓練データの多様性の不足などにより単純すぎるモデルを作ってしまう可能性、そしてその結果生じる、訓練データでもテストデータでも性能が落ちる適合不足 (underfitting) を理解している。
4. 交差検証法 (cross validation (leave-one-out)) を理解している。例えば、3 次式、2 次式、1 次式による回帰モデルのうち、相対的に優れたモデルを、交差検証法による予測平均二乗誤差の大きさによって選ぶことができる。

9. 因果推論

因果関係とは、ある 2 つの事象が原因と結果という関係にあることをいい、相関とは異なる概念である。様々な変数について因果関係を明らかにすることにより、様々な課題解決を図れる。

9.1 ★★因果モデル

1. 異なる変数に同時に影響を与える交絡因子の考え方を説明できる。
2. 回帰モデルにおける偏相関係数を理解している。
3. 条件付き独立性の概念（特に多変量正規分布に従う確率変数族の場合）を理解している。
4. 傾向スコアの定義と基本性質（バランシング性など）を説明できる。

9.2 ★★★グラフィカルモデリング

1. 多変量正規分布の共分散行列の推定法を理解している。
2. 非巡回的有向独立グラフの確率構造を、数式および図で説明できる。
3. ソフトウェアを用いてグラフィカルモデルの推定を実行できる。

10. 深層学習・ニューラルネットワーク

非線形構造を考慮したモデルの多くは、その構造を事前に想定しなければならない。しかし、深層学習では柔軟な非線形構造を実現できるため、多くの場面で有益な結果を残している。また、学習データやテストデータの作成さえも仕組みに組み込むことで、これまでに得られなかったモデルの構築を可能とし、AI のさらなる発展につながっている。

10.1 ニューラルネットワークの仕組み

1. ★★単純および多層パーセプトロンの入力・出力構造を視覚的に理解している。
2. ★★非線形活性化関数によって複雑な非線形関係へ対応可能となることを理解している。
3. ★★誤差逆伝播法による多層パーセプトロンの学習原理を理解している。
4. ★★★畳み込みニューラルネットワークにおける畳み込み層（特徴抽出）とプーリング層（相対位置情報の平滑化・頑健化など）による学習行程を理解している。
5. ★★★畳み込みニューラルネットワークについて、画像解析など実際の応用例を挙げることができる。
6. ★★★言語、音声、信号など様々な時系列データを扱うための RNN (Recurrent Neural Network), LSTM (Long and short-term memory), GRU (Gated recurrent unit) などの考え方を説明できる。

10.2 ★★★深層生成モデル

1. (確率的) 生成モデルの意義およびその主要な推定方法を理解している。
2. 深層ニューラルネットワークを通じて様々な非線形構造を持つデータを生成できることを説明できる。
3. VAE (variational autoencoder) や GAN (Generative adversarial networks) といった画像生成モデルなど、深層生成モデルの具体例いくつかを原理も含め説明できる。

高等学校学習指導要領とスキルセット及び学修目標の対応

※ 下線は中分類との対応を示す。

数学（現行学習指導要領（平成 21 年告示））

数学 I

数と式

数と集合 式 計算基礎：情報、コンピュータの仕組み

図形と計量

三角比 図形の計量

二次関数

二次関数とそのグラフ 二次関数の値の変化 数学基礎：微積分

データの分析

データの散らばり データの相関 データの記述・可視化：データの記述 データの記述・可視化：データの可視化

数学 II

いろいろな式

式と証明 高次方程式

図形と方程式

直線と円 軌跡と領域

指数関数・対数関数

指数関数 対数関数 数学基礎：微積分

三角関数

角の拡張 三角関数 三角関数の加法定理 数学基礎：微積分

微分・積分の考え

微分の考え 積分の考え 数学基礎：微積分

数学 III

平面上の曲線と複素数平面

平面上の曲線 複素数平面

極限

数列とその極限 関数とその極限 数学基礎：微積分 数学基礎：数列

微分法

導関数 導関数の応用 数学基礎：微積分

積分法

不定積分と定積分 積分の応用 数学基礎：微積分

数学 A

場合の数と確率

場合の数 確率 統計基礎：確率と確率分布

整数の性質

約数と倍数 ユークリッドの互除法 整数の性質の活用

図形の性質

平面図形 空間図形

数学 B

確率分布と統計的な推測

確率分布 正規分布 統計的な推測 統計基礎：確率と確率分布 統計基礎：統計的推測

数列

数列とその和 漸化式と数学的帰納法 数学基礎：数列

ベクトル

平面上のベクトル 空間座標とベクトル 数学基礎：線形代数

情報（次期学習指導要領（平成 30 年告示））

情報 I

情報社会の問題解決

- ・ 情報やメディアの特性、情報と情報技術を活用した問題の発見・解決
- ・ 情報に関する法規や制度、情報セキュリティ、情報社会における個人の責任及び情報モラル データの法規・倫理：情報倫理、情報セキュリティ データの法規・倫理：データに関する法律・規制
- ・ 情報技術が人や社会に果たす役割と及ぼす影響

コミュニケーションと情報デザイン

- ・ メディアの特性とコミュニケーション手段 計算基礎：情報、コンピュータの仕組み
- ・ 情報デザインが人や社会に果たしている役割
- ・ 効果的なコミュニケーションを行うための情報デザイン

コンピュータとプログラミング

- ・ コンピュータや外部装置の仕組みや特徴、コンピュータでの情報の内部表現と計算に関する限界 計算基礎：情報、コンピュータの仕組み
- ・ アルゴリズムを表現する手段、プログラミングによるコンピュータや情報通信ネットワークの活用 計算基礎：アルゴリズムとプログラミング
- ・ 事象のモデル化、シミュレーションを通じたモデル評価と改善

情報通信ネットワークとデータの活用

- ・ 情報通信ネットワークの仕組みや構成要素、プロトコルの役割及び情報セキュリティを確保するための方法や技術
- ・ データを蓄積、管理、提供する方法、情報通信ネットワークを介して情報システムがサービスを提供する仕組みと特徴 データの取得・管理・加工：データ管理とデータ形式
- ・ データを表現、蓄積するための表し方、データを収集、整理、分析する方法 データの記述・可視化：データの記述 データの記述・可視化：データの可視化

第 1 次報告のスキルセット 初級レベル（コア☆のみ）

データサイエンス教育を全国の全ての大学に展開するにあたり、各大学の学生の特性や文系・理系の違いを考慮する必要がある。このため概ね 1.5 単位分程度と考えられる内容を特にコアとして推奨した。さらに、データサイエンスを学ぶ意義、重要性の理解や学修の動機付け等を図るために、7 分野に加えて「データサイエンスを学ぶ意義」を設けた。各大学の実情に応じて参照されたい。なお、各大学において「データサイエンス入門」のような科目を設置する場合には、「データサイエンスを学ぶ意義」及びコアに各大学の事情に応じて内容を追加することで、2 単位や 4 単位の講義を設計できるなど、一定の自由度を持つものであると考えている。

大分類	中分類	小分類	概要	キーワード	レベル
データサイエンスを学ぶ意義	(データサイエンスを学ぶ意義、重要性の理解) 1. データ駆動型社会 2. データに基づく課題解決、意思決定の支援 3. データサイエンスのサイクル 4. データサイエンスの様々な事例				☆
データの法規・倫理	2. データに関連する法律・規制	2.1 個人のデータに関連する法規	個人情報保護法の概要	個人情報保護法、改正個人情報保護法（第三者提供、個人情報の定義の拡充）	☆
			個人のデータに関連する国際的な動き	EU 一般データ保護規則 (GDPR)、忘れられる権利	☆
		2.2 統計法	統計法の概要	統計法の基本事項と意義、基幹統計調査	☆
	3. データサイエンスの倫理	3.1 倫理に配慮したデータ収集と利活用	データ収集対象者からの同意の必要性、オプトアウトの仕組み	インフォームドコンセント、オプトアウト	☆
		3.2 データの匿名化	匿名化と仮名化の違い、個人の再識別の可能性	匿名化、仮名化、再識別、秘密の曝露や差別の誘引	☆
		3.3 データサイエンスに関する様々なバイアス	データサイエンスに関するバイアス	データバイアス、アノテーションバイアス、標本選択バイアス、帰納バイアス	☆
データの記述・可視化	1. データの記述	1.1 種々のデータ	データの分類	質的変数と量的変数、名義尺度、順序尺度、間隔尺度、比例尺度	☆
			時系列データ	時系列データ、季節調整、移動平均	☆
		1.2 基本統計量	分布の中心とばらつきに関する基本統計量	最頻値、中央値、平均値、四分位範囲、分散と標準偏差	☆
			分布における相対的な位置の概念	標準化、標準得点（ z 得点）、偏差値、外れ値	☆
		1.3 相関関係	2 つの量的変数の相関関係	散布図、共分散、相関係数	☆
			多変数データ（3 つ以上の量的変数）の 2 変数間の相関関係	散布図行列、相関係数行列	☆
			2 つ以上の質的変数の相関関係	分割表、クロス集計表、リスク比、リスク差、オッズ、オッズ比	☆
			相関関係と因果関係の違い	相関と因果、交絡、偏相関係数	☆
			説明変数が 1 つの単回帰分析	最小二乗法、回帰分析、説明変数、目的変数、予測値、残差、決定係数	☆
	2. データの可視化	2.1 グラフの構成要素	グラフの構成要素	色の効果や特徴、点の色・大きさ・形状への配慮、線の太さと様々な破線	☆
		2.2 統計グラフ	データ全体の傾向や特徴を読み取るためのグラフ	棒グラフ、折れ線グラフ、円グラフ、帯グラフ	☆
			グラフの作り方、見方のリテラシー	チャートジャンク、不必要な視覚的要素	☆
		2.3 分布の統計グラフ	ヒストグラム、箱ひげ図、散布図などのデータの分布を示すグラフ	ヒストグラム、箱ひげ図、散布図（2 変数）、ヒートマップ（3 変数以上）	☆

データの取得・管理・加工	1. データ取得とオープンデータ	1.1 日本や世界のオープンデータ	日本政府のオープンデータの取り組みとオープンデータの意義	二次利用可能なルール，機械判読への適性，国，地方公共団体，民間事業者の取り組み	☆
			オープンデータの普及に向けた国際的な動き	2013 年の G8 サミットで合意された「オープンデータ憲章」	☆
		1.2 オープンデータの取得	統計ポータルサイトからのデータのダウンロード	e-Stat, DATA.GO.JP データカタログサイト, Open Knowledge International	☆
			メタデータの検討	メタデータ，データ取得の経緯，変数の特徴	☆
	2 データ管理とデータ形式	2.1 表形式のデータ	表形式のデータ	csv, tsv	☆
統計基礎	1. 確率と確率分布	1.1 場合の数，順列・組み合わせ	場合の数，順列・組み合わせ	場合の数，順列・組み合わせ，階乗	☆
		1.2 確率	確率の概念．加法定理などの基本的な定理．ベイズの定理	加法定理，条件付き確率，ベイズの定理	☆
		1.3 確率分布，確率変数	確率変数や確率分布の概念．離散型・連続型の確率分布	確率変数，確率関数，確率密度関数，累積分布関数	☆
			確率変数の平均（期待値）と分散	平均，期待値，分散	☆
		1.4 主要な確率分布	複数の確率変数に関する諸概念	同時分布，周辺分布，条件付分布，共分散と相関，独立	☆
			ベルヌーイ分布，二項分布	ベルヌーイ分布，二項分布，期待値と分散	☆
			正規分布	正規分布．期待値と分散．線形変換による正規性の保存	☆
			その他の主要な 1 変量確率分布（1）	ポアソン分布，指数分布，一様分布	☆
	2. データ収集法と確率構造	2.1 標本調査	2 変量正規分布の視覚的理解	2 変量正規分布，相関	☆
			標本を抽出して調査し，その結果から母集団の性質を推測する標本調査	単純無作為抽出法，クラスター抽出法，2 段階抽出法	☆
			統計量を手がかりとした母集団の特性の推測	母数と統計量（母平均と標本平均）の区別，標本分布	☆
数学基礎	1. 線形代数	1.1 ベクトル	ベクトル	n 個の数値の組としての n 次元ベクトル ($n=2, 3, \dots$)	☆
			ベクトルの基本的な演算	ベクトルの和，スカラー倍，線形結合	☆
			内積	内積，ベクトルの直交性，ノルム，単位ベクトル	☆
	2. 微積分	2.1 多項式関数	二次関数，多項式と基本性質	二次関数や多項式のグラフ，四則演算	☆
			二項定理と二項係数	二項定理，パスカルの三角形	☆
	3. 数列	3.1 データと数列	データの標本平均や分散の表現	標本平均，加重平均，標本分散，ベクトルの内積	☆

第 1 次報告のスキルセット 初級レベル（☆及び★）

大分類	中分類	小分類	概要	キーワード	レベル
データサイエンスを学ぶ意義	(データサイエンスを学ぶ意義、重要性の理解) 1. データ駆動型社会 2. データに基づく課題解決、意思決定の支援 3. データサイエンスのサイクル 4. データサイエンスの様々な事例				☆
データの法規・倫理	1. 情報倫理、情報セキュリティ	1.1 情報倫理・関連法規	情報倫理と関連法規の基礎	関連する法律（知的財産、個人情報の保護、不正アクセス行為の禁止等）	★
			情報社会で生活する上でのマナー、モラル、倫理の意義	マナー、モラル、倫理	★
		1.2 情報セキュリティ	情報セキュリティの 3 要素	情報セキュリティの 3 要素 機密性・完全性・可用性	★
			情報セキュリティを確保するための対策	パスワード、個人認証、ソフトウェアのセキュリティ更新プログラム	★
	2. データに関連する法律・規制	2.1 個人のデータに関連する法規	個人情報保護法の概要	個人情報保護法、改正個人情報保護法（第三者提供、個人情報の定義の拡充）	☆
			個人のデータに関連する国際的な動き	EU一般データ保護規則 (GDPR)、忘れられる権利	☆
		2.2 統計法	統計法の概要	統計法の基本事項と意義、基幹統計調査	☆
	3. データサイエンスの倫理	3.1 倫理に配慮したデータ収集と利活用	データ収集対象者からの同意の必要性、オプトアウトの仕組み	インフォームドコンセント、オプトアウト	☆
		3.2 データの匿名化	匿名化と仮名化の違い、個人の再識別の可能性	匿名化、仮名化、再識別、秘密の曝露や差別の誘引	☆
		3.3 データサイエンスに関する様々なバイアス	データサイエンスに関するバイアス	データバイアス、アンノテーションバイアス、標本選択バイアス、帰納バイアス	☆
		3.4 公表バイアス	公表バイアス、出版バイアス	公表バイアスの危険性、公表バイアスを減らす取り組み	★
データの記述・可視化	1. データの記述	1.1 種々のデータ	データの分類	質的変数と量的変数、名義尺度、順序尺度、間隔尺度、比例尺度	☆
			時系列データ	時系列データ、季節調整、移動平均	☆
		1.2 基本統計量	分布の中心とばらつきに関する基本統計量	最頻値、中央値、平均値、四分位範囲、分散と標準偏差	☆
			分布における相対的な位置の概念	標準化、標準得点（ z 得点）、偏差値、外れ値	☆
		1.3 相関関係	2 つの量的変数の相関関係	散布図、共分散、相関係数	☆
			多変数データ（3 つ以上の量的変数）の 2 変数間の相関関係	散布図行列、相関係数行列	☆
			2 つ以上の質的変数の相関関係	分割表、クロス集計表、リスク比、リスク差、オッズ、オッズ比	☆
			相関関係と因果関係の違い	相関と因果、交絡、偏相関係数	☆
			説明変数が 1 つの単回帰分析	最小二乗法、回帰分析、説明変数、目的変数、予測値、残差、決定係数	☆
			説明変数が 1 つの単回帰分析	説明変数が 1 つの単回帰分析	☆
	2. データの可視化	2.1 グラフの構成要素	グラフの構成要素	色の効果や特徴、点の色・大きさ・形状への配慮、線の太さや様々な破線	☆
		2.2 統計グラフ	データ全体の傾向や特徴を読み取るためのグラフ	棒グラフ、折れ線グラフ、円グラフ、帯グラフ	☆
			グラフの作り方、見方のリテラシー	チャートジャンク、不必要な視覚的要素	☆
		2.3 分布の統計グラフ	ヒストグラム、箱ひげ図、散布図などのデータの分布を示すグラフ	ヒストグラム、箱ひげ図、散布図（2 変数）、ヒートマップ（3 変数以上）	☆
データの取得・管理・加工	1. データ取得とオープンデータ	1.1 日本や世界のオープンデータ	日本政府のオープンデータの取り組みとオープンデータの意義	二次利用可能なルール、機械判読への適性、国、地方公共団体、民間事業者の取り組み	☆
			オープンデータの普及に向けた国際的な動き	2013 年の G8 サミットで合意された「オープンデータ憲章」	☆
		1.2 オープンデータの取得	統計ポータルサイトからのデータのダウンロード	e-Stat、DATA.GO.JP データカタログサイト、Open Knowledge International	☆
			メタデータの検討	メタデータ、データ取得の経緯、変数の特徴	☆
	2. データ管理とデータ形式	2.1 表形式のデータ	表形式のデータ	csv、tsv	☆
		2.2 その他のデータ形式	データ記述言語、SNS のつながりや鉄道路線などのデータ	XML、JSON、離散グラフ、キー・バリュ形式である隣接リスト	★
		2.3 データベース	データベースの意義と運用・管理	データベース管理システム (DBMS)	★

			リレーショナルデータベース	正規化, 選択, 射影, 結合, SQL	★
			キー・バリュー型に対応したデータベース	NoSQL, キー・バリュー型に対応したデータベース	★
	3. データの前処理	3.1 データクレンジング	データクレンジング	文字列の表記の揺れの吸収 (文字列, 数字, 日付, 時刻)	★
		3.2 外れ値, 異常値, 欠損値	外れ値, 異常値, 欠損値	外れ値, 異常値 (箱ひげ図, 標準化得点), データ欠損の要因, 欠損値の補完	★
		3.3 データ加工	データ加工	部分集合の抽出, 行の並び替え, 新しい列の追加	★
統計基礎	1. 確率と確率分布	1.1 場合の数, 順列・組み合わせ	場合の数, 順列・組み合わせ	場合の数, 順列・組み合わせ, 階乗	☆
		1.2 確率	確率の概念. 加法定理などの基本的な定理. ベイズの定理	加法定理, 条件付き確率, ベイズの定理	☆
		1.3 確率分布, 確率変数	確率変数や確率分布の概念. 離散型・連続型の確率分布	確率変数, 確率関数, 確率密度関数, 累積分布関数	☆
			確率変数の平均 (期待値) と分散	平均, 期待値, 分散	☆
			複数の確率変数に関する諸概念	同時分布, 周辺分布, 条件付分布, 共分散と相関, 独立	☆
		1.4 主要な確率分布	ベルヌーイ分布, 二項分布	ベルヌーイ分布, 二項分布, 期待値と分散	☆
			正規分布	正規分布. 期待値と分散. 線形変換による正規性の保存	☆
			その他の主要な 1 変量確率分布 (1)	ポアソン分布, 指数分布, 一様分布	☆
			2 変量正規分布の視覚的理解	2 変量正規分布, 相関	☆
	2. データ収集法と確率構造	2.1 標本調査	標本を抽出して調査し, その結果から母集団の性質を推測する標本調査	単純無作為抽出法, クラスタ抽出法, 2 段抽出法	☆
			統計量を手がかりとした母集団の特性の推測	母数と統計量 (母平均と標本平均) の区別, 標本分布	☆
		2.2 ランダム化比較試験	ランダム化比較試験, 根拠に基づく医療. 根拠に基づく政策立案. A/B テスト	実験研究と観察研究, ランダム化比較試験. A/B テスト	★
	3. 統計的推測	3.1 統計的モデル	母集団に対して複数の確率分布の候補を表現する統計的モデル, 分布をコントロールする母数 (パラメータ)	統計的モデル, 母数, パラメータ	★
		3.2 標本分布	独立同一分布 (i. i. d.) に従う確率変数の概念. その標本平均の期待値と分散, 標本分散の期待値	独立同一分布, 標本平均, 標本分散	★
		3.3 点推定, 区間推定	点推定. 特にモーメント法, 最尤法による推定量の構成	モーメント法, 最尤法	★
			推定量の性質. 特にバイアス, 不偏性, 平均二乗誤差	バイアス, 不偏推定量, 平均二乗誤差, バイアス分散分解	★
			信頼区間の構成. 信頼係数の理解	信頼区間と信頼係数, 平均の区間推定, 比率の区間推定	★
		3.4 仮説検定の概念	二項分布の成功確率の検定などの基本的な仮説検定を通じた仮説検定のロジック	帰無仮説, 対立仮説, 2 種の誤り	★
			有意水準, 検出力, p 値など検定に関する指標. 帰無仮説のもとで検定統計量の従う分布	有意水準, 検出力, p 値, 検定統計量	★
数学基礎	1. 線形代数	1.1 ベクトル	ベクトル	n 個の数値の組としての n 次元ベクトル ($n=2, 3, \dots$)	☆
			ベクトルの基本的な演算	ベクトルの和, スカラー倍, 線形結合	☆
			内積	内積, ベクトルの直交性, ノルム, 単位ベクトル	☆
		1.2 行列	行列の定義と基本的な演算	行列の和とスカラー倍	★
			行列の積	行列の積. 積の非可換性. 結合則, 分配則	★
	2. 微積分	2.1 多項式関数	二次関数, 多項式と基本性質	二次関数や多項式のグラフ, 四則演算	☆
			二項定理と二項係数	二項定理, パスカルの三角形	☆
		2.2 初等関数	指数関数, 対数関数と基本性質	指数関数の形の把握, 指数法則. 対数関数の形の把握, 対数法則	★
			三角関数と基本性質	三角関数の形の把握, 三角関数の公式	★
			自然対数の底と指数関数	自然対数の底, ネイピア数	★
			関数の極限の定義と計算	関数の極限	★
			導関数と関数の増減	平均変化率, 微分係数と接線. 導関数	★
		2.3 1 変数関数の微分法	初等関数の導関数	初等関数の導関数	★
			微分法の基本的な関係	合成関数の微分法, 積の微分法	★
			導関数と関数の極値と最大・最小	関数の最大・最小, 極大・極小と導関数	★
			微分の逆演算と原始関数	原始関数, 不定積分	★
		2.4 1 変数関数の積分法	定積分の計算	定積分と微分積分学の基本定理	★
			積分の基本的な関係式	部分積分, 置換積分	★
		2.5 2 変数関数の	偏微分の定義	偏微分, 偏導関数, 接平面	★

		微分法	2 変数関数の最大最小問題	最大最小問題，極値と偏導関数の関係，ヘッセ行列	★	
		2.6 2 変数関数の積分法	長方形領域での重積分	重積分，重積分（長方形領域），累次積分	★	
	3. 数列	3.1 データと数列	データの標本平均や分散の表現	標本平均，加重平均，標本分散，ベクトルの内積	☆	
		3.2 数列	基本的な数列	等比数列，等差数列	★	
	4. 演習		数列および数列の和の極限	数列の極限，数列の和の極限	★	
		4.1 線形代数演習	ベクトルや行列を対象とした計算機演習	R や Python を用いた計算機演習	★	
		4.2 微積分演習	微積分に関する計算機演習	Excel, R, Python などの既存の関数，ライブラリを用いた数値計算数式処理の演習（Python の Sympy パッケージなど）	★	
	計算基礎	1. 情報，コンピュータの仕組み	1.1 数と表現	2 進数の表現，基数変換の方法，負の数の表現，2 進数の加算や減算，表現可能な数値の範囲など	2 進数，2 の補数表現	★
			1.2 有効数字，計算の誤差	コンピュータが計算処理を有限の桁数で行うことから生じる誤差．浮動小数点演算と誤差．丸め誤差	有効数字，仮数部，指数部	★
			1.3 データ量の単位	ビット，バイトなどのデータ量の単位や，SI 接頭辞を使った表現	ビット，バイトなどの情報量，SI 接頭辞(k, M, G, T, P, m, μ , n, p など)	★
1.4 デジタル化			アナログとデジタルの特徴，デジタル化のプロセス，画像データ	量子化，標本化，符号化，画像データ（ラスタ，ベクタ）	★	
1.5 文字の表現			コンピュータの内部では，文字を数値で表現していることの理解．文字コード	ASCII, Unicode	★	
1.6 集合と命題			集合と命題，ベン図，真理値表などの基本的な考え方	真偽値，否定，論理和，論理積	★	
1.7 論理演算			論理演算の考え方と基本的な演算，及び真理値表の利用方法	論理回路，論理ゲート，AND, OR, NOT, NAND, XOR	★	
2. データ構造		2.1 配列，リスト	サイズが既知の複数の要素（値）の集合を格納・管理する配列，サイズが不明のデータを柔軟に格納するリスト	メモリ，ベクトル，行列，アドレス	★	
		2.2 連想配列	項目（キー）と値（バリュー）を組とするデータを格納するデータ構造	連想，辞書，ハッシュ	★	
3. アルゴリズムとプログラミング		3.1 アルゴリズム	フローチャートの記号と処理手順の表現方法，疑似コード，アクティビティ図	端子，処理，判断，矢印	★	
			アルゴリズムの基本構造である順次構造，選択構造（分岐構造），繰返し構造（ループ構造）	代入，計算，順次構造，選択構造，繰返し構造	★	
			合計，探索，併合（マージ），整列（ソート）などの基本アルゴリズム	バブルソート，幅優先探索，深さ優先探索	★	
		3.2 プログラミング	Python や R などのインタプリタ型言語	ソースコード，解釈実行，対話的実行環境	★	
			プログラムの書き方の規則である構文と，構文要素の解釈を定める意味論の区別	変数，代入，文，計算	★	
	データ型，扱うデータを格納するためその種類に応じた変数の宣言		文字型，整数型，浮動小数点型	★		
	演算子と式．それぞれのデータ型に対して適用できる演算の種類とその表記法		四則演算，論理演算，式	★		
	繰り返し処理や場合に応じた処理を記述する制御構造		for, while, if	★		
	計算処理単位をまとめる関数．その定義と呼出し		関数，引数，返値，再帰呼出し	★		
	ファイルや標準入出力などに対するデータの読み書き		標準入出力，ターミナル，ファイル，print, scanf	★		
	プログラムの誤り（バグ）の有無を検証するデバッグ，テスト		デバッグ，テスト	★		
	ツールやライブラリを用いて具体的な問題を実際に解く応用プログラミング		表計算，テキスト処理，Excel，ライブラリ，API	★		
モデリングと評価	1. 教師あり学習	1.1 回帰分析	回帰分析の手続き	★		
		1.2 ロジスティック回帰	ロジスティック回帰の手続き	★		

	2. 教師なし学習	2.1 クラスタリング(k -平均法)	k -平均法の手続き	k -平均法の手続き, 結果の頑健性(初期値, 距離), 統計ソフトウェアによる実データを用いたデータ解析	★
		2.2 階層クラスタリング	階層クラスタリングの手続き	階層クラスタリングの手続き, 樹形図, 結果の頑健性(類似度, 併合法), 統計ソフトウェアによる実データを用いたデータ解析	★
	3. モデルの評価	3.1 評価指標	回帰問題における平均二乗誤差	平均二乗誤差	★
			二値分類問題における混同行列に基づく評価	正解率(accuracy), 適合率(precision), 再現率(recall), 混同行列	★
		3.2 訓練データとテストデータ	訓練データとテストデータ	訓練データとテストデータ, 汎化誤差, 過学習, 適合不足	★
			交差検証法	交差検証法, leave-one-out	★

数理・データサイエンス教育強化拠点コンソーシアムの概要

どの大学・どの学部に進学しても、全ての学生が今後必要となる数理的思考力とデータ分析・活用能力を体系的に身に付けることが出来る環境の構築を目指す。

■ コンソーシアムの沿革

- 2016 年度 大学の数理・データサイエンス教育の強化方策（文部科学省）
拠点校選定（6 大学：北海道大学、滋賀大学、大阪大学、東京大学、京都大学、九州大学）
- 2017 年度 コンソーシアム設立
- 2019 年度 協力校選定（20 大学）、全国展開に向けた 6 ブロック化
- 2020 年度 協力校（3 大学）・特定分野協力校（7 大学）選定
連携校の公募を開始（2021 年 4 月現在 84 大学、2 短大、国立高等専門学校機構）

■ コンソーシアムの主な役割

- 全国的なモデルとなる標準カリキュラム・教材を協働して作成するとともに、他大学への普及方策（例えば全国的なシンポジウムの開催等）の検討・実施
- 各大学のセンターにおける教育内容・教育方法の好事例を共有し、より取組を進展させるための議論を行うなど、センターの情報交換等を行うための対話の場の設定

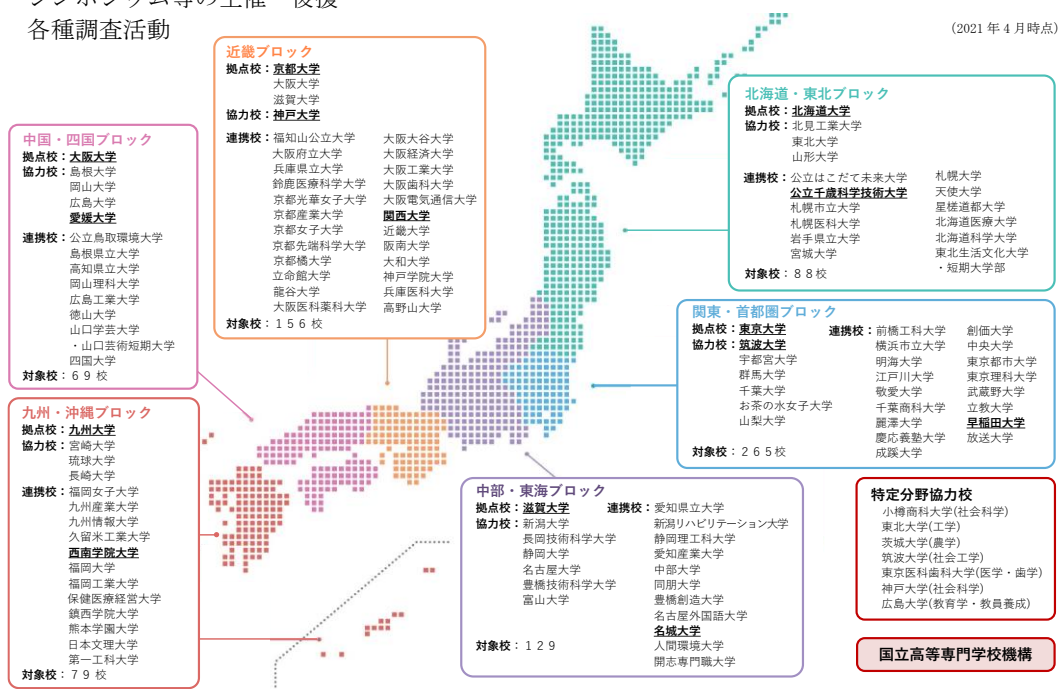
■ 3 分科会及び特別委員会

カリキュラム分科会
・標準カリキュラムの作成
教材分科会
・教科書シリーズの企画編集 ・講義動画・スライド等の収集・公開
教育用データベース分科会
・教育用各種データの収集・公開 ・教育に活用可能な実データ・課題の収集・公開

モデルカリキュラム（リテラシーレベル） の全国展開に関する特別委員会
・モデルカリキュラム（リテラシーレベル）の検討（2020 年 4 月公表）
モデルカリキュラム（応用基礎レベル） の全国展開に関する特別委員会
・モデルカリキュラム（応用基礎レベル）の検討（2021 年 3 月公表）

■ その他の活動

- 情報発信・活動紹介（ホームページ、ニュースレター、SNS）
- シンポジウム等の主催・後援
- 各種調査活動



※下線の大学は拠点校又はブロック別の代表校を示す。これらの大学でコンソーシアム運営会議を構成。

数理・データサイエンス教育強化拠点コンソーシアム カリキュラム分科会

委 員

主 査	河合 玲一郎	東京大学大学院総合文化研究科／数理・情報教育研究センター教授
副主査	林 和則	京都大学高等教育院附属データ科学イノベーション教育研究センター教授
	行木 孝夫	北海道大学大学院理学研究院教授
	姫野 哲人	滋賀大学データサイエンス学部准教授
	高野 渉	大阪大学数理・データ科学教育研究センター特任教授
	増田 弘毅	九州大学大学院数理学研究院教授

第2次報告 審議経過

(2020 年)

- 6 月 22 日 第 4 回 カリキュラム分科会
 - ・ 第 2 次報告の方向性の検討、役割分担の決定
- 7 月 30 日 第 5 回 カリキュラム分科会
 - ・ 第 2 次報告の方針確認、スキルセットの議論
- 8 月 26 日 第 6 回 カリキュラム分科会
 - ・ 第 2 次報告スキルセットの検討
- 10 月 22 日～ コンソーシアム協力校（特定分野協力校含む）
- 11 月 9 日 全 30 大学への意見照会
- 12 月 25 日 第 7 回 カリキュラム分科会
 - ・ 第 2 次報告学修目標の方針確認

(2021 年)

- 1 月 26 日 第 8 回 カリキュラム分科会
 - ・ 第 2 次報告学修目標の検討
- 2 月 29 日 第 9 回 カリキュラム分科会
 - ・ 第 2 次報告学修目標の検討
- 3 月 16 日 第 10 回 カリキュラム分科会
 - ・ 第 2 次報告素案の検討
- 5 月 21 日 第 11 回 カリキュラム分科会
 - ・ 第 2 次報告素案の検討
- 4 月 26 日～ コンソーシアム拠点校、代表校、協力校（特定分野協力校含む）
- 5 月 24 日 への意見照会