

## 数理・データサイエンス・AI 教育プログラム（リテラシーレベル） 全学共通科目「データサイエンス A」講義内容（全15回）

### 1. オリエンテーション：ビッグデータ時代のデータサイエンス・AI

ビッグデータやデータサイエンス、AI を取り巻く過去数十年にわたる技術的变化を理解し、産業界の動向、消費生活の変化といった社会的状況について概観する。また、ChatGPT をはじめとした生成 AI についても触れ、現在大きな技術的変革の中にあることを理解する。その他、本講義の進め方について説明する。

### 2. 概要・手法：記述統計（1）

数値データに関する1変数の主な記述統計量の意味や導出方法を理解することと、Excelなどで記述統計量や棒グラフなどのグラフを作成できることを狙いとしている。

### 3. 概要・手法：記述統計（2）

2変数の記述統計については、元データの形が(i) 質的変数×質的変数・(ii) 質的変数×量的変数・(iii) 量的変数×量的変数の3パターンあることを示す。特に(iii)については、標本共分散と標本相関係数の意味や導出方法を理解することと、Excelなどで散布図などのグラフを作成することも狙いとしている。

### 4. 概要・手法：データサイエンスに関連する諸技術の概要

データとは何か（構造化データ、非構造化データの違い）、データベース、プログラミング言語、ニューラルネットワークなど、基盤となる技術についてキーワードレベルで理解する。また、データを使った分析手法としての機械学習、最適化、情報検索、自然言語処理、ロボット技術、画像・映像処理について、キーワードレベルで理解する。

### 5. ケーススタディ：健康情報と疫学 - エビデンスを作る手法 -

医学や健康科学の分野における意思決定にはエビデンス（科学的根拠）が不可欠である。もちろん、サービス受益者（患者や国民など）の preference（嗜好）などの価値観やサービス提供者（医療者など）の熟練性や国の経済環境など様々な因子がこの意思決定には関係するが、エビデンスは特に重要である。この講義では、エビデンスを作る手法として、医学や健康科学の分野において発展してきた epidemiology（疫学）の歴史を紐解き、データを判断や意思決定に生かす視点を学ぶ。原因が不明でも、予防対策を立てられる疫学の特長にも触れる。

### 6. ケーススタディ：スポーツにおけるデータサイエンスの活用

スポーツ選手のトレーニングや競技力向上にデータサイエンスがいかに関用されているか。1. ゲーム分析、2. フィジカルフィットネス測定、3. 動作を3本柱とした事例紹介を通じて、理解を深めるとともに将来的な可能性について考察する。

### 7. ケーススタディ：産業連関分析と国と国のつながり

世界各国は SNA（国民経済計算）の一つとして産業連関表というデータを公表している。貿易統計を利用して各国の産業連関表を接続した国際産業連関表がある。本講義では、日本と中国の国際産業連関表を用いて、国と国が経済的にどのようにつながっているかイメージできることを目的とする。

8. **ケーススタディ：統計データとアンケートを用いた需要予測**  
生産計画や販売計画を立案する場合，何らかの根拠に基づいた需要予測値がないと計画の立案ができない．ひとつの方法に，もとなる信頼性の高い数値データと利用者や消費者がどのような行動をするのかが予測できるデータを用いて需要を予測する方法がある．本講義では，物流量の予測を公表されている統計データを用いて実際に需要予測された手法を理解し，予測値が出来上がるまでの過程を理解することを目的とする．
9. **ケーススタディ：ことばのデータサイエンス**  
日本語を例にして，言語を対象にしたデータサイエンスとはどういうものか理解する．文章のジャンルや内容，書き手によって変化する文章中の特徴とはどのようなものか理解する．また，用意されたデータをもとに，簡単な分析が自分でもできるようになること．
10. **ケーススタディ：人文情報学におけるデータサイエンス：オープンな漢字データベース**  
データサイエンスは社会科学への応用として注目されることが多いが，自然言語処理の分野でも長年研究されており，今般の生成 AI のブームによって急激な展開が起きている．一方で人文情報学の分野においてはデジタルアーカイブといった情報処理技術が発展してきたが，扱う資料の多義性によりまとまりがなく，発展途上の段階にあるともいえる．講義では，人文学におけるデータサイエンスの基盤となる文字処理技術の例としてオープンな漢字データベースを取り上げ，その背景と現状を紹介することで，特に人文系学部の学生にとってもデータサイエンス分野についての興味を持ってもらうことを目標とする．
11. **ケーススタディ：デジタル・ヒューマニティーズの実践**  
文学，歴史，哲学といった伝統的な学問に代表される人文学（ヒューマニティーズ）は，従来，テキストの精読を基礎に発展してきた．しかし，近年，人文学の分野においてもデジタル技術を応用した研究がさかんになされている．本学でも，書道資料や漢籍のデジタル・アーカイブを公開するなどの取り組みをしている．本講義では，人文学の分野で行われているデジタル技術の活用事例を紹介し，文系の学生でも手軽に活用できるデジタル・ツールの実例を紹介する．
12. **ケーススタディ：巨大な数の列挙と組み合わせ最適化**  
組み合わせ爆発とは何か，簡単そうに見える問題でもサイズが少し増えるだけで最新の計算機でも全く歯が立たないこと，また，そのように爆発する計算量を如何にサボるかということこそが AI の本質であることを理解する．
13. **ケーススタディ：データサイエンスと社会シミュレーション**  
データを分析して人や社会の因果関係を見つけるだけでは必ずしも現実の課題解決に結び付くとは限らない．解決策を見出してもそれがどんな結果をもたらすのかを検証し意思決定する必要がある．本講義では，この様な科学的な意思決定が，情報御術の飛躍的な進歩により可能になりつつあることを示す．
14. **概要・手法：データリテラシーとデータ・AI 利活用における留意事項**  
この授業全体は，データサイエンス・AI についてすべての学生が身に付けるべき“デジタル社会の基礎知識”を習得するためのものであり，所謂「リテラシーレベル」の内容となっている．しかし，データ・AI におけるバイアス特性やプライバシー保護・セキュリティに関わる課題等の理解，およびデータクリーニングやデータクレンジングといった，データを扱う上での留意事項は，次のレベル（「応用基礎レベル」）に進むためにも是非必要な事柄である．本講義の狙いは，それらの留意事項の意味を理解し，実社会においてデータ・AI を扱う場合に心がけておくべき事柄の認識を持ってもらうことである．
15. **まとめ：データ・AI 利活用に関するまとめ / データ・AI 利活用の最新動向と今後の課題**  
本講義は，現実社会におけるデータ・AI 利活用の動向・実態を概観しつつ，科目全体のまとめを行

う. 講義は2つの部分から構成される. 第1の部分は, 過去30年ほどのデジタル技術の進歩が経済・社会に与えた影響について, 俯瞰(マクロ)的な視点から捉える. まず, IT(情報通信)関連産業で続出した革新的なビジネス・モデルを概観し, 次いでIoT(Internet of Things)に代表される, より広範な産業でのデジタル変革の動向を展望する. いずれの産業においても, データ・AIの利活用がどのように新しい価値を生んでいるかについて, 具体例を中心に把握していく. 第2の部分は, ビジネスの現場を念頭に, データ・AIの利活用の実態を当事者(ミクロ)的な視点から捉える. データ・AI利活用の際の標準的手順や注意点についても触れる.