

Course number		U-LAS11 10007 LJ55					
Course title (and course title in English)		データ分析基礎 Basic Data Analysis		Instructor's name, job title, and department of affiliation		Institute for Liberal Arts and Sciences Program-Specific Senior Lecturer,SEKIDO HIROTO Graduate School of Informatics Program-Specific Associate Professor,SATO HIROYUKI	
Group		Natural Sciences		Field(Classification)		Data Science(Foundations)	
Language of instruction		Japanese		Old group		Group B	
Number of weekly time blocks		1		Class style		Lecture (Face-to-face course)	
Days and periods		Tue.5		Target year		All students	
				Eligible students		For all majors	
[Overview and purpose of the course]							
<p>本授業の目的は、データ分析の基本的な事項を学び、実際に大規模データ解析を行うための基礎的な技術を身につけることにある。なお、大規模データ解析や最新のデータ分析手法を行うにあたっては、既存のソフトウェアを用いて簡潔にできるとは限らない。そこで、本授業では、データ解析の基礎となる線形代数と最適化手法についても講義する。ただし、厳密な数学的証明は必要最小限に抑え、直感的な理解を深める事を目標とする。</p> <p>はじめに、データを扱う上で必要不可欠な確率変数・平均・分散などの統計の基礎的な概念について解説する。また、仮説検定の考え方を述べ、いくつかの例を用いて解説する。</p> <p>その後、複数の変数の関係性を調べる多変量解析を行う際に必要となる線形代数について講義する。特に、連立一次方程式の解法であるガウスの消去法とLU分解、回帰分析において必要となるQR分解、主成分分析などで重要となる固有値分解・特異値分解について解説する。</p> <p>次に、多変量解析の基礎である回帰分析と主成分分析について講義する。また、回帰分析、主成分分析について理解を深めるためExcelと統計ソフトRを使って実習を行う。</p> <p>最後に、スパース推定を題材に、最適化手法についての講義と実習を行う。</p>							
[Course objectives]							
<p>1. 確率変数・平均・分散など統計の基礎的な内容、および、仮説検定について理解する。</p> <p>2. 多変量解析の基礎である回帰分析、主成分分析についての概念を理解し、データ分析に応用できるようになる。</p> <p>3. Excelと統計ソフトRを使いこなす知識を身につけ、実際に簡単なデータ解析を行えるようになる。</p> <p>4. データ分析に登場する最適化問題とそれを解くアルゴリズムについて、基本的な考え方を理解する。</p>							
[Course schedule and contents]							
<p>授業回数はフィードバックを含め全15回とする。</p> <p>・ガイダンス、統計の基礎（関戸，1回） 授業の概要について説明する。その後、確率変数・平均・分散など統計の基礎的な内容について講義する。</p> <p>・仮説検定（関戸，1回） 仮説検定について、その基本的な考え方について講義する。</p>							
-----							
Continue to データ分析基礎(2)							

## データ分析基礎(2)

### ・線形代数の基礎（關戸，2回）

行列、ベクトルについての定義からはじめ、連立一次方程式の解法であるガウスの消去法やLU分解、回帰分析において必要となるQR分解、主成分分析などで重要となる固有値分解と特異値分解について解説する。

### ・回帰分析（關戸，2～3回）

予測などに使われる回帰分析について講義する。主にExcelを用いて実習も行い、実習を通じて、ダミー変数の扱い方、多重共線性の問題と解決法、モデル選択の考え方の基本、Ridge回帰とLasso回帰などを教授する。

### ・主成分分析（關戸，2～3回）

多変量解析の基礎である主成分分析について講義する。回帰分析と組み合わせる主成分回帰などについても解説する。時間があれば、低ランク近似の考えに基づく、EMアルゴリズムを用いた欠損値の推定法とその応用などについても述べる。主に統計ソフトRを用いて実習を行う。

### ・最適化の基礎とスパース推定のための最適化手法（佐藤，5回）

いろいろな種類の最適化問題の紹介からはじめ、連続最適化問題を解くための基礎的な手法である最急降下法について講義する。また、Lasso回帰などのスパース推定法において現れる最適化問題を効率的に解くアルゴリズムを教授し、データ分析への応用について述べるとともに、実習を行う。

### ・フィードバック（關戸・佐藤，1回）

## [Course requirements]

None

## [Evaluation methods and policy]

定期試験、レポート、平常点を総合的に評価する。

### ・定期試験（配点50点）

講義で解説した基本的な原理や理論を理解できているかどうかを評価する。

### ・レポート評価（配点25点）

授業の途中に、講義で解説した「線形代数の基礎知識と統計処理の基礎的な内容」、「最適化の基礎的な内容」についてそれぞれレポート課題を出題する。

### ・平常点（配点25点）

出席状況と質問など通した授業への積極的な参加を評価する。

## [Textbooks]

教科書は使用しない。

資料が必要な場合には、授業中に配布する。

## [References, etc.]

### （References, etc.）

Introduced during class

Continue to データ分析基礎(3)

## データ分析基礎(3)

### [Study outside of class (preparation and review)]

行列については授業中にも解説するが、行列の扱いに不慣れな学生は予習あるいは復習することが望ましい。

また、厳密な数学的議論など授業中に省略した事項について、各自で学習することを期待する。

### [Other information (office hours, etc.)]

オフィスアワーについては担当教員の KULASIS 登録情報を参照すること。

關戸啓人: [sekido@amp.i.kyoto-u.ac.jp](mailto:sekido@amp.i.kyoto-u.ac.jp)

佐藤寛之: [hsato@amp.i.kyoto-u.ac.jp](mailto:hsato@amp.i.kyoto-u.ac.jp)

授業時間外で質問がある場合には、あらかじめ、上記のアドレスにメールをすること。