



京都大学  
KYOTO UNIVERSITY



全学共通教育

# ビッグデータの時代 データ科学を学ぼう

国際高等教育院附属

データ科学イノベーション教育研究センター 山本章博

2021年4月

# その裏側には、 データサイエンス。

お手元に  
届いていると  
思います

今話題のデータサイエンスは、京都大学全学共通教育で学ぶことができます

ともだちの  
始まりの裏側にも



健康維持の  
裏側にも



最高の買い物の  
裏側にも



あの宅配の  
裏側にも



# 内閣府・文部科学省はデータ科学(データサイエンス)の教育強化を進めています

デジタル社会の「**読み・書き・そろばん**」である「**数理・データサイエンス・AI**」の基礎などの必要な力を**全ての国民**が育み、あらゆる分野で人材が活躍

## 主な取組

## 育成目標【2025年】

### エキスパート

#### 先鋭的な人材を発掘・伸ばす環境整備

- 若手の自由な研究と海外挑戦の機会を拡充
- 実課題をAIで発見・解決する学習中心の課題解決型AI人材育成

### 応用基礎

#### AI応用力の習得

- AI×専門分野のダブルメジャーの促進
- AIで地域課題等の解決ができる人材育成(産学連携)

#### 認定制度・資格の活用

- 大学等の優れた教育プログラムを政府が認定する制度構築
- 国家試験(ITパスポート)の見直し

### リテラシー

#### 学習内容の強化

- 大学の標準カリキュラムの開発と展開
- 高校におけるAIの基礎となる実習授業の充実

#### 小中高校における教育環境の整備

- 多様なICT人材の登用(高校は1校に1人以上、小中校は4校に1人以上)
- 生徒一人一人が端末を持つICT環境整備

トップクラスの育成  
100人程度/年

2,000人/年

25万人/年  
(高校の一部、  
高専・大学の50%)

50万人/年  
(大学・高専卒業生**全員**)

100万人/年  
(高校卒業生**全員**)  
(小中学生**全員**)

すべての大学卒業生が  
修得しているべき素養

# 数理・データサイエンス教育強化拠点コンソーシアム

## 近畿ブロック

拠点校：京都大学

大阪大学 滋賀大学

協力校：神戸大学

特定分野協力校：神戸大学(社会科学)

連携校：京都府立大学 福知山公立大学 大阪府立大学  
兵庫県立大学 鈴鹿医療科学大学 京都光華女子大学  
京都産業大学 京都女子大学 京都先端科学大学  
京都橘大学 京都ノートルダム女子大学  
同志社大学 立命館大学 龍谷大学  
大阪医科薬科大学 大阪大谷大学 大阪経済大学  
大阪工業大学 大阪電気通信大学 関西大学  
近畿大学 四天王寺大学・四天王寺大学短期大学部  
阪南大学 桃山学院大学 大和大学  
神戸学院大学 兵庫大学 兵庫医科大学  
高山山大学

対象校：155校

## 中国・四国ブロック

拠点校：大阪大学

協力校：鳥根大学 岡山大学 広島大学 愛媛大学

特定分野協力校：広島大学(教育学・教員養成)

連携校：香川大学 高知大学 公立鳥取環境大学  
鳥根県立大学 高知県立大学 岡山理科大学  
広島工業大学 徳山大学  
山口学芸大学・山口芸術短期大学 四国大学

対象校：68校

## 九州・沖縄ブロック

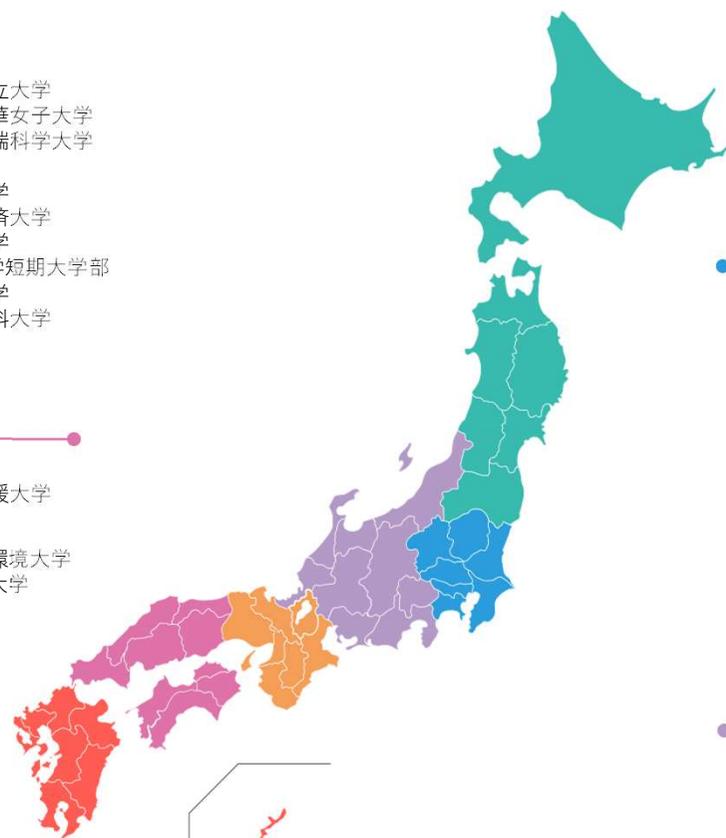
拠点校：九州大学

協力校：宮崎大学 琉球大学 長崎大学

連携校：福岡女子大学 九州産業大学  
九州情報大学 久留米工業大学  
西南学院大学 日本経済大学  
福岡大学 福岡工業大学  
保健医療経営大学 活水女子大学 鎮西学院大学  
長崎総合科学大学 熊本学園大学 日本文理大学  
第一工業大学

対象校：79校

京都大学は拠点校  
R4年度より近畿ブロック代表校



## 北海道・東北ブロック

拠点校：北海道大学

協力校：北見工業大学 東北大学 山形大学

特定分野協力校：小樽商科大学(社会科学) 東北大学(工学)

連携校：公立ほこだて未来大学 公立千歳科学技術大学  
札幌市立大学 札幌医科大学 岩手県立大学  
札幌大学 天使大学 星槎道都大学  
北海道医療大学 北海道科学大学 稚内北星学園大学  
東北工業大学 東北生活文化大学・短期大学部

対象校：89校

## 関東・首都圏ブロック

拠点校：東京大学

協力校：筑波大学 宇都宮大学 群馬大学  
千葉大学 お茶の水女子大学 山梨大学

特定分野協力校：茨城大学(農学) 筑波大学(社会工学)  
東京医科歯科大学(医学・歯学)

連携校：前橋工科大学 横浜市立大学 明海大学  
江戸川大学 敬愛大学 千葉商科大学  
麗澤大学 東海大学 慶応義塾大学  
工学院大学 国士館大学 上智大学  
成蹊大学 成城大学 創価大学  
中央大学 東京医療保健大学 東京経済大学  
東京都市大学 東京理科大学 法政大学  
武蔵野大学 立教大学 早稲田大学  
放送大学

対象校：263校

## 中部・東海ブロック

拠点校：滋賀大学

協力校：新潟大学 長岡技術科学大学 静岡大学  
名古屋大学 豊橋技術科学大学 富山大学

連携校：愛知県立大学 新潟リハビリテーション大学  
静岡理工科大学 愛知産業大学 中京大学  
中部大学 同朋大学 豊橋創造大学  
名古屋外国語大学 名城大学 人間環境大学  
開志専門職大学

対象校：128校

<http://www.mi.u-tokyo.ac.jp/consortium/pdf/newsletter09.pdf>

# 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～

## ■ 背景

政府の「AI戦略2019」（2019年6月策定）にて、リテラシー教育として、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を習得する、とされたことを踏まえ、各大学・高専にて参照可能な「モデルカリキュラム」を数理・データサイエンス教育強化拠点コンソーシアムにおいて検討・策定。

## ■ 学修目標・カリキュラム実施にあたっての基本的考え方

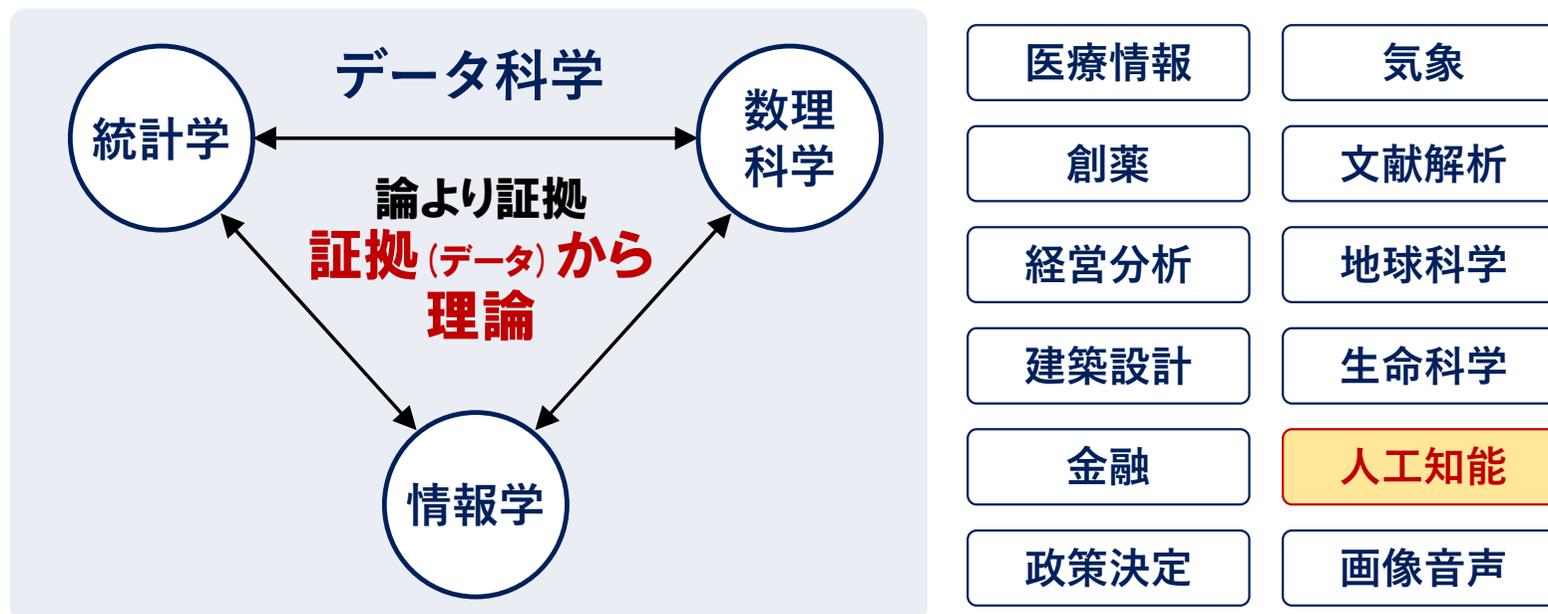
今後のデジタル社会において、数理・データサイエンス・AIを**日常生活、仕事等の場で使いこなすことができる基礎的素養**を主体的に身に付けること。そして、学修した数理・データサイエンス・AIに関する知識・技能をもとに、これらを扱う際には、**人間中心の適切な判断ができ、不安なく自らの意志でAI等の恩恵を享受し、これらを説明し、活用できるようになること。**

1. 数理・データサイエンス・AIを活用することの「**楽しさ**」や「**学ぶことの意義**」を重点的に教え、学生に好奇心や関心を高く持ってもらう魅力的かつ特色ある教育を行う。数理・データサイエンス・AIを活用することが「好き」な人材を育成し、それが自分・他人を含めて、次の学修への意欲、動機付けになるような「**学びの相乗効果**」を生み出すことを狙う。
2. 各大学・高専においてカリキュラムを実施するにあたっては、各大学・高専の教育目的、分野の特性、個々の学生の学習歴や習熟度合い等に応じて、本モデルカリキュラムのなかから適切かつ柔軟に**選択・抽出し教育を行う**。
3. **実データ、実課題を用いた演習**など、**社会での実例を題材**に数理・データサイエンス・AIを活用することを通じ、現実の課題と適切な活用法を学ぶことをカリキュラムに取り入れる。
4. リテラシーレベルの教育では、統計数理を駆使した「**厳密さ**」を専ら追求するのではなく、むしろ「**分かりやすさ**」を重視する。  
[http://www.mi.u-tokyo.ac.jp/consortium/pdf/model\\_literacy.pdf](http://www.mi.u-tokyo.ac.jp/consortium/pdf/model_literacy.pdf)

# データ科学（データサイエンス）って何？

データを収集してコンピュータで管理し、  
数的手法を用いて分析することにより結論を導き、将来の推測を行う学問

## 情報学・統計学・数学（数理科学）の融合分野



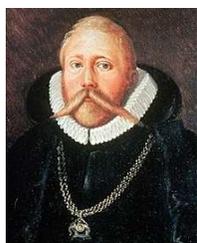
データ科学イノベーション教育研究センターは全学共通教育での  
統計学を中心としたデータ科学科目を整備・設計・実施する組織です

# データと物理法則

物理法則の源泉はデータである

## 16世紀頃まで惑星の運行は謎だった

- 恒星は一定の動き：同じ時刻に観測すると位置が少しずつずれていく
- 惑星はときどき逆行するなど単純ではない動きをする



Tycho Brahe  
(1546-1601)<sup>\*1</sup>



Johannes Kepler  
(1571-1630)<sup>\*2</sup>

ティコ・ブラーエによる  
火星の運行に関する**高精度観測データ**

ヨハネス・ケプラーの発想  
“惑星の軌道は楕円ではないか?”

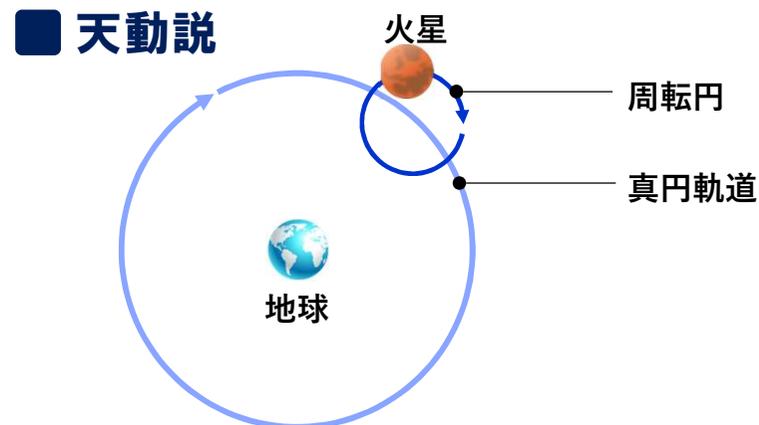
地動説の下で、火星の軌道が楕円だと仮定すると  
**データが説明できることが分かった!**

**ケプラーの法則 (第1法則)**

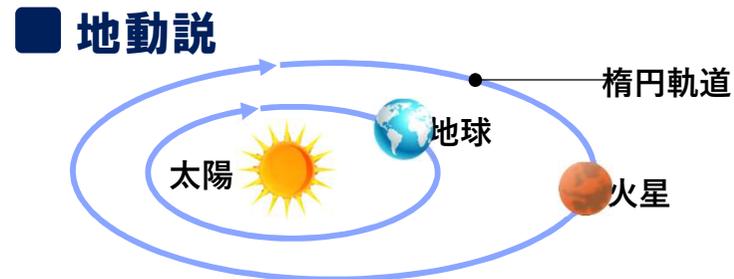
惑星の軌道は楕円である!

さらにニュートンの万有引力の法則へと発展

### ■ 天動説



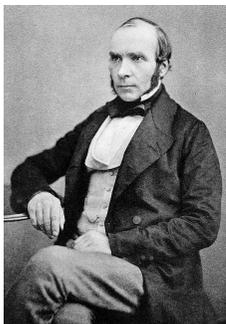
### ■ 地動説



- 天動説の場合は、真円軌道に小さな周転円を追加するモデル
- 地動説の場合は、真円軌道だと誤差が大きい

# データ科学と医療・政策

## 1854年ロンドンにおいて特定の井戸周辺にコレラ多発



John Snow  
(1813-1858) \*3

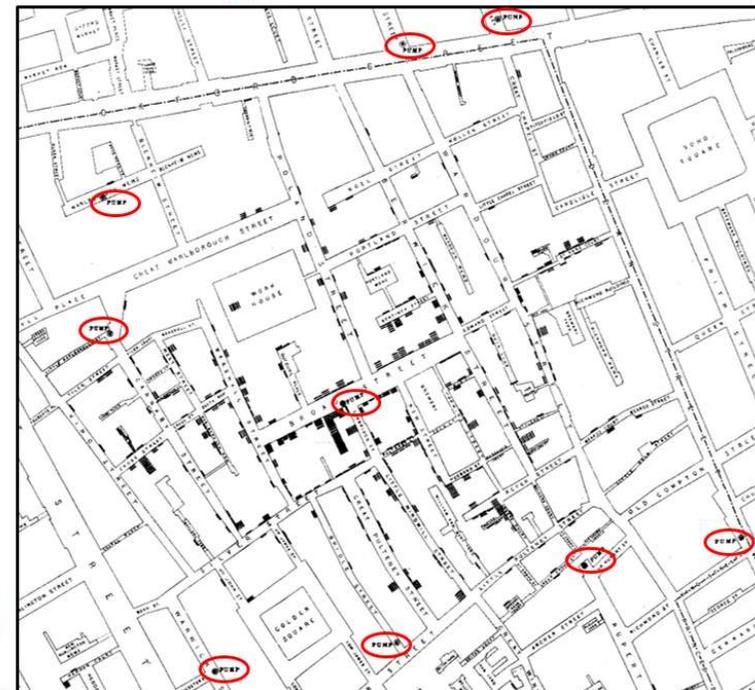
データを集めることにより同一の区域でも、水道供給会社によってコレラ発生率に大きな違いがあることを発見

### コレラ対策という政策が進む

#### 流行ピーク、8割がインフルエンザの外来 インフルエンザ

検査	あり	なし	計	
陽性	1440	8 偽陽性	1448	陽性的中率 99.4%
陰性	160 偽陰性	392	552	陰性的中率 71.0%
計	1600	400	2000	
	感度 90%	特異度 98%		

赤枠がポンプの場所



検査結果が陰性でも感染している確率は、全体の感染率と感度から計算される

#### 数学

$$P_B(A) = \frac{P_A(B) P(A)}{P(B)}$$

# データ科学と人工知能(AI)

2011年衆議院会議録作成システムに京都大学で開発された音声認識技術が採用



従来



衆議院本会議

\*5

新システム



事務事業の仕分けをはじめとした  
ご提案をいただきました。

情報学

膨大な量の国会質疑音声記録と国会議事録  
のデータベース

統計学

音声記録と議事録の違いを数式化

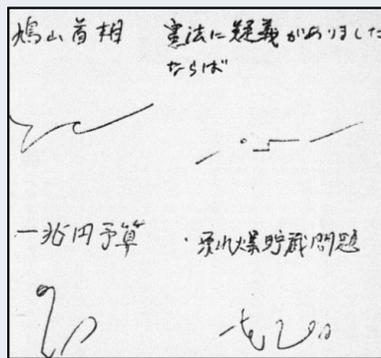
数学

$$P_B(A) = \frac{P_A(B) P(A)}{P(B)}$$

$$a_n = \frac{1}{\pi} \int_{-\pi}^{\pi} f(t) \cos nt dt, (n = 0, 1, 2, 3, \dots)$$



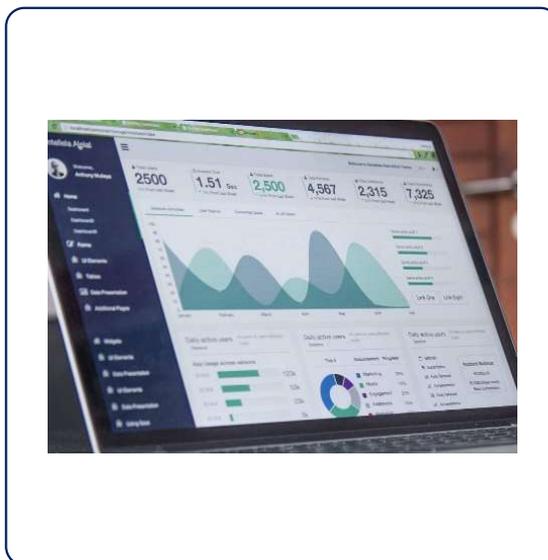
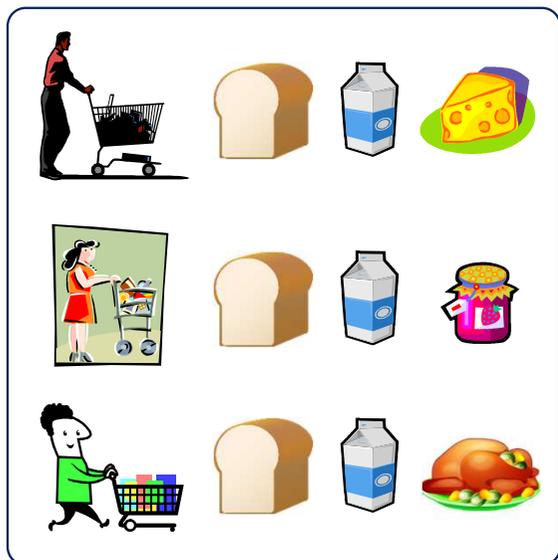
\*6



\*7

# データ科学とビジネス

## ■ お店の販売データ（POSデータ）分析



### 販売・仕入れ・商品開発などにデータを活用

- 一緒に売れる商品の組み合わせは？
- 最近よく売れている/売れ残りがちな商品は？
- これから売れそうな商品は？

## データ科学分野・情報群科目の例

	データ科学分野 (自然群)	情報群
<b>リテラシーレベル</b> (全回生向け) 情報・統計に関する 一般常識・基本知識	<ul style="list-style-type: none"><li>■ 数理・データ科学のための数学入門Ⅰ・Ⅱ</li><li>■ 統計入門</li><li>■ 数理統計</li></ul>	<ul style="list-style-type: none"><li>■ 情報基礎</li><li>■ 情報基礎演習</li><li>■ 情報と社会</li><li>■ 情報基礎実践</li><li>■ プログラミング演習 (Pythonなど)</li></ul>
 <b>専門基礎レベル</b> (主に2回生以上向け) データから分布などの 情報を抽出する力	<ul style="list-style-type: none"><li>■ 統計と人工知能</li><li>■ データ分析演習Ⅰ・Ⅱ</li><li>■ データ分析基礎</li></ul>	<ul style="list-style-type: none"><li>■ 情報企業論</li><li>■ 情報ネットワーク</li></ul>

データサイエンススクール (課外)

# 統計入門

## ■ 全学部・全回生向（医学部医学科1回生は必須）

※学部・学科によっては「数理統計」が指定・推奨されているので「履修の手引き」「履修要覧」などをよく確認してください。

## ■ 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に対応

## ■ 合格者は修了証を受領可能

## ■ より広範な学生が統計学の基本的な考え方を理解することを目指す

## ■ 数学的な厳密な理論を理解することよりも、データを扱うエンドユーザーとしてデータの性質に応じた適切な分析方法を選択できるようになることを目指す

## ■ 統計解析ソフトを自習形式で取り入れる

## ■ 「統計入門」では扱いきれなかった話題（分散分析，回帰分析，因果推論など）を「統計と人工知能」で講述



CIREDS

データ科学イノベーション  
教育研究センター



MDASH  
Literacy

修了証番号：1234567890

修 了 証

生年月日 2022年01月01日

氏 名 mdash事務局

上記の者は、下記の科目を修得し、本学の「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）」を修了したことを認めここに証します。

記

令和3年度 「統計入門」

以上

令和4年3月31日

京都大学国際高等教育院長 宮川 恒



# データ分析基礎 + データ分析演習 I, II

## データ分析基礎

- 大規模データ解析を行うための基礎的な技術を習得
- 既存のソフトウェアを用いて簡潔に実行できない場合のために、どのように計算されているのかについても触れる

例

ガウスの消去法, QR分解, 固有値分解・特異値分解

- 厳密な数学的証明は必要最小限、直感的な理解を深める
- 回帰分析, 主成分分析, Excelと統計ソフトRを使った実習

## データ分析演習 I, II

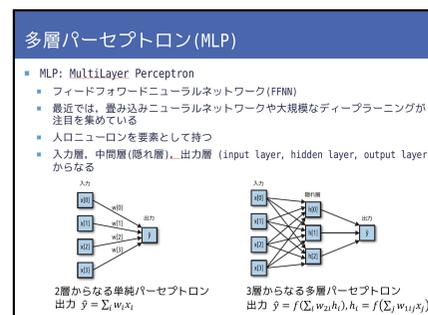
- 大規模なデータ分析の演習を通じてデータ分析の基礎の体得を目標
- データ分析手法や統計解析ソフトは、履修者が今後専門とする分野や興味に合わせて幅広い選択ができるように、異なる内容の演習科目を揃える

例1

気象データとPython言語を用いた回帰分析, 機械学習, クラスタリング

例2

医学・医療データとExcel, R言語を用いた統計処理 (統計検定2~3級レベル)



## データ科学の学び方

### 現代の「読み・書き・算盤」

▶ 誰でも将来必要となるときが来ます。

各学部・学科の履修方針に従って情報学・統計学・数学(数理科学)をバランスよく学んでください

- 全学共通教育で学習しなければならない内容は、学部・学科によって異なります。

### 全学共通教育での統計学関連科目についての詳細は

データ科学イノベーション教育センターが提供する科目一覧をご覧ください

データサイエンスを学ぶためのガイダンス

<https://ds.k.kyoto-u.ac.jp/guidance/>



## 科目を履修できない場合のために

**履修計画は必ず各学部・学科の履修方針に従って立ててください。**

- 単位数や時間割などの制約によって、データ科学に関する科目を希望通りに履修できない可能性もあります。

**データ科学センターではデータ科学に関する自習用の教材や企画を用意しています。**

協力：高等教育研究開発推進センター，京大オリジナル(株)

- MOOC (e-learning)
- プログラミング自習キット
- データサイエンス入門講座
- データサイエンス・スクール

# MOOC「統計の入門」

## 学内e-learningプラットフォームKoALAで公開

協力：高等教育研究開発推進センター  
<https://koala.highedu.kyoto-u.ac.jp/>



### 差の差の分析法

- 「商品購入増へのテレビCMの効果測定する」
  - CMが無くても増えたのではないのか？
  - CMを「見た」から「購入意向が高くなった」のか？
  - もともと「購入意向が高い」からCMを「見た（覚えやすい）」のか？

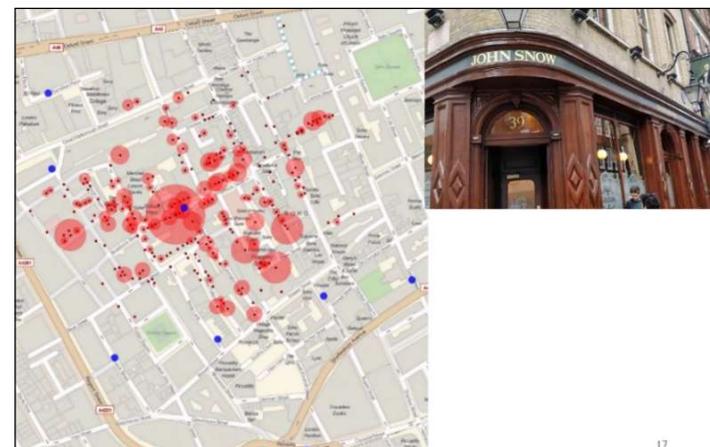
効果測定の特徴

CMの効果については、CMに触れることで高まった購入意向から、CMに触れなくても高まった購入意向を引いた値で評価

グループ	出稿前 (%)	出稿後 (%)	増減分 (%)
テレビCM接触 (効果測定対象群)	21.4%	24.9%	+3.5%
テレビCM非接触 (コントロール群)	17.4%	17.8%	+0.4%

<https://www.is.nri.co.jp/report/opinion/2017/000200.html>

- 因果関係を説明するための手法として「差の差の分析法」も有効な手法の一つ



- 全学共通科目「統計入門」の要点を習得できます
- 「統計入門」の予習に利用することもできます

# Pythonプログラミング自習キット

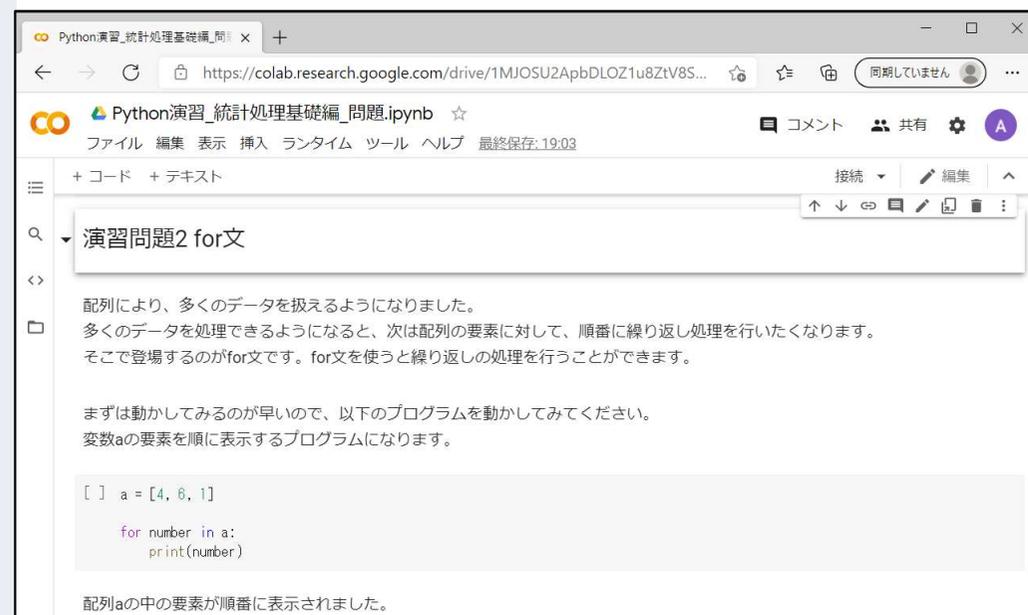
[https://kubar.rd.iimc.kyoto-u.ac.jp/intro\\_python/](https://kubar.rd.iimc.kyoto-u.ac.jp/intro_python/)

データ科学で注目されている  
**プログラミング言語Python**  
の初歩を習得するためのキットです

Internet環境とWebブラウザだけを用いて  
学習できます。

■Google社が提供するプログラミング環境  
Google Colaboratoryを利用しますので、  
Googleの個人アカウントが必要です

■京都大学の学生だけが利用できます



# データサイエンススクール(随時開催)



2019 於 京都大学 吉田キャンパス 吉田南構内

School 25 8.6 (火) 10:30 - 18:00 / School 26・27 8.29・30 (木・金) 10:30 - 18:00

データ科学は科学研究の基本の一つであり、本学における数多くの研究活動において活用されています。このデータサイエンススクールは「数理及びデータサイエンスに係る教育強化」拠点大学の活動として、多くの学生がデータ科学の最前線について知る事ができ、さらにデータ科学の手法について、その利用法を体験しながら実践的に学ぶことができる機会を提供するように企画いたしました。

2018 年より、データ科学教育研究センターでは各種スクールを開催し、多くの方に好評いただきました。今回の Data Science School 2019-August-においても、講師の先生方をお招きし、データ科学手法についてご教授いただきます。演習の時間では受講生が講義で学んだ手法を実データとコンピュータを用いて実践するようになっています。

データ科学を学習・体験できるまたとない機会ですので是非ご参加ください。

School 25 8/6 10:30 - 18:00 学術情報メディアセンター南館 203 号室 定員 50 名

## 「文系のための SPSS によるデータ分析入門」

京都大学教育学研究科 准教授 岡邊 健

本セミナーは、統計解析ソフト SPSS を用いたデータ分析の入門講座です。SPSS の使用経験がない方に向けて、操作の基本をご紹介します。論文(卒論など)の執筆に向けてデータ分析にチャレンジしてみたい文系学生が主たる対象です。統計学の基礎的知識の有無は問いません。

対象: 本学学部学生・大学院生・研究者 \*本セミナーは SPSS のライセンスが本学蔵書(2022 年度)からの利用は行っていません。

School 26・27 8/29・30 10:30 - 18:00 学術情報メディアセンター南館 301 号室 定員 50 名

## 「Python によるデータ分析」

株式会社セラク DX 本部 データサイエンス事業部

### 農業データの活用の実際」

主査 金井 伸也、島崎 信二

日本における農業は生産人口の減少や高齢者の低下などの課題を抱えており、ICT 技術による生産の効率化が求められている。そのためには、日々の生産環境の分析をすることで、適切な栽培管理を行うことが重要である。

本講座では、プログラミング言語 Python を用いて、圃場(はじょう)の時系列データの分析手法を紹介する。講義はハンズオン形式で、実際に手を動かすことに重きを置き、得られた結果に対する議論も行う。

準備: Python 実行環境として Web ブラウザから Google Colaboratory を利用する。

事前に Google のアカウント作成と、Colaboratory の使用方法の確認をしておくこととスムーズである。

対象: 本学学部学生・大学院生・研究者、数理・データサイエンス教育強化拠点コンソーシアム、データ関連人材育成関西地区コンソーシアム、HeKKSaGOn 参加の大学の学生・研究者

## アクセスマップ:

京都大学 吉田南構内  
学術情報メディア  
センター南館 301



## お申し込み方法:

ホームページの専用フォームからのお申し込みとなります。  
願書情報をアップいたします。  
ぜひ参加登録ください。



## 参加費: 無料

※本学に在籍する学生が対象です。  
※本学に在籍しない学生は、内容が変更になる場合があります。ご了承ください。  
※Web 申し込みは SPSS のライセンスが本学蔵書(2022 年度)からの利用は行っていません。 \*本セミナーは SPSS のライセンスが本学蔵書(2022 年度)からの利用は行っていません。

<http://ds.k.kyoto-u.ac.jp/school/m201908/>

## 主催

京都大学国際高等教育院 附属データ科学イノベーション教育研究センター

contact@ds.k.kyoto-u.ac.jp

詳細は

データ科学イノベーション教育研究センターホームページ

<http://ds.k.kyoto-u.ac.jp/> で確認してください。

## 開講科目だけではカバーできない内容を補います

- 実践的 (演習的) 内容
- 先端的内容 (機械学習・AI など)

## 春期休業, 夏期休業, 週末などを利用します

## ■ 1日~4日間の集中講義

## 学内の研究科・研究所・附属病院所属, 他大学, および企業から講師を招聘して開講します

- 先端研究で用いられる手法を  
計算機演習を通じて体得

# データサイエンス入門講座（有料）

主催：京大オリジナル（株）

共催：京都大学データ科学イノベーション教育研究センター



講義動画(15分程度×18回)  
+1回のライブ配信  
(質疑応答と「統計検定」の模擬試験)

「統計検定」を実施する  
一般財団法人統計質保証推進協会との  
連携セミナー

■ 「統計検定」の過去問を利用します

レベルに合わせて学習可能

■ 文系学生向3級, 2級, 統計ソフトR

詳細は京大オリジナル（株）ホームページ、  
データ科学イノベーション教育研究センターで  
確認してください。

## おわりに

### 01 データ科学は現代の「読み・書き・算盤」です

- 誰でも必要となる 때가 来ます

### 02 データ科学を学ぶには、各学部・学科の履修方針に従って 統計学・情報学・数学(数理科学)をバランスよく学んでください

- 各学部・学科で学ばなければならない基礎科目や専門科目をおろそかにしないでください。

### 03 データ科学イノベーション教育研究センターは、 皆さんのデータ科学の学習をサポートします

- 全学共通教育の科目を担当しています
- MOOC, 自習キット, データサイエンススクール, データサイエンス入門講座を提供しています
  - ・ データ科学科目を履修できない場合に活用してください

## 図版等出典(最終確認日:2022年3月日)

---

- \*1 [https://commons.wikimedia.org/wiki/File:Tycho\\_Brahe.JPG](https://commons.wikimedia.org/wiki/File:Tycho_Brahe.JPG)
  - \*2 <https://commons.wikimedia.org/wiki/File:Kepler.png>
  - \*3 [https://commons.wikimedia.org/wiki/File:John\\_Snow.jpg](https://commons.wikimedia.org/wiki/File:John_Snow.jpg)
  - \*4 [https://en.wikipedia.org/wiki/1854\\_Broad\\_Street\\_cholera\\_outbreak#/media/File:Snow-cholera-map-1.jpg](https://en.wikipedia.org/wiki/1854_Broad_Street_cholera_outbreak#/media/File:Snow-cholera-map-1.jpg)
  - \*5 [http://www.shugiin.go.jp/internet/index.nsf/html/images/gijidou002.jpg/\\$File/gijidou002.jpg](http://www.shugiin.go.jp/internet/index.nsf/html/images/gijidou002.jpg/$File/gijidou002.jpg)
  - \*6 猿谷 豊, 衆議院における音声認識を利用した会議録作成業務,情報管理  
Journal of Information Processing and Management : Vol. 55 (2012) No.6 p.392-399  
(<https://doi.org/10.1241/johokanri.55.E1>)
  - \*7 衆議院式速記の一例  
[https://commons.wikimedia.org/wiki/File:Shugiin-system\\_Shorthand\\_01.jpg](https://commons.wikimedia.org/wiki/File:Shugiin-system_Shorthand_01.jpg)
-