



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

公式Twitter [[@cireds2020](#)]

2023年4月

新入生の皆様, ご入学おめでとうございます. この時間は, 「ビッグデータの時代
データ科学を学ぼう」と題しまして
全学共通教育でのデータ科学について説明いたします
私はデータ科学イノベーション教育研究センターの山本です.

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



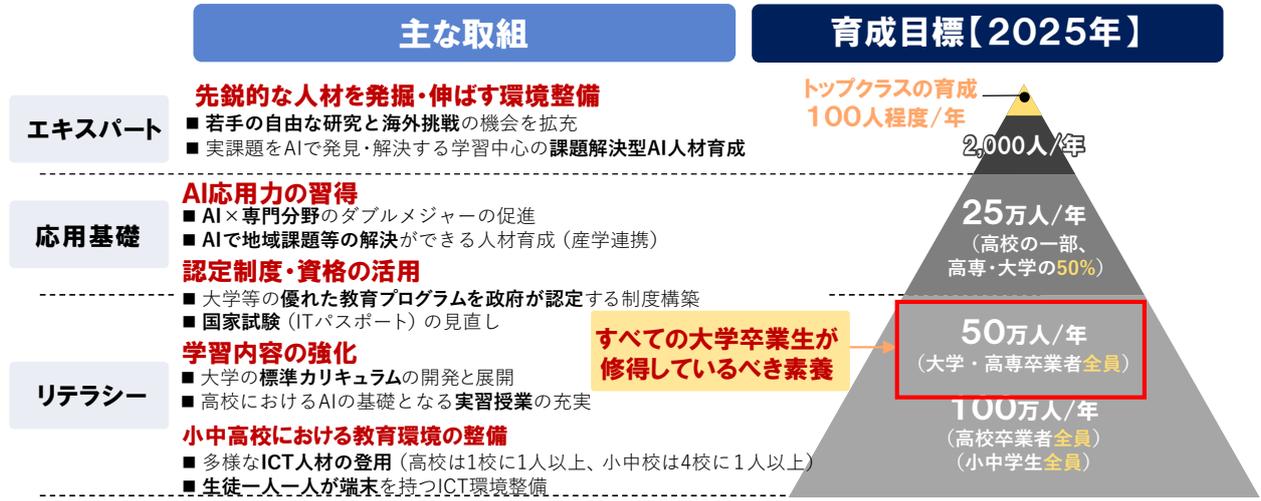
皆さんが合格と同時に受け取られた宅配便の中には、このような案内を入れさせていただきました。

手に取っていただいて、これは何のことを言っているのだろう、と疑問に思われた方もいらっしゃると思います

この案内は皆さんに少しでもデータ科学を学んでいただく作成しました。

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

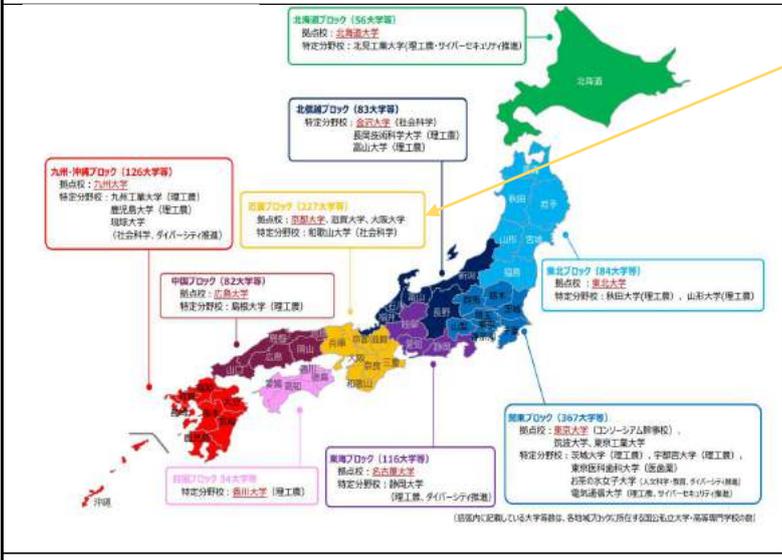


現在、内閣府と文部科学省は、大学におけるデータサイエンスあるいはデータ科学の教育の普及を強く推進しています。

データサイエンスの基礎的内容は、すべての大学卒業生が修得しているべき、基本的な素養としています。

データサイエンスは、デジタル社会の読み・書き・算盤と位置付けられ、情報化社会において必須の素養と認識されているからです。

数理・データサイエンス教育強化拠点コンソーシアム



京都大学は拠点校 R4年度より近畿ブロック代表校

拠点校: 滋賀大学, 京都大学, 大阪大学
特定分野協力校: 和歌山大学 (社会科学)
連携校: [国立] 三重大学, 京都工芸繊維大学, 神戸大学
 [公立] 京都府立大学, 福知山公立大学, 大阪公立大学, 兵庫県立大学
 [私立] 皇學館大学, 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, 京都精華大学, 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, 同志社女子大学, 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, 大阪歯科大学, 大阪成蹊大学, 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, 摂南大学, 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, 畿央大学, 高野山大学, 大阪キリスト教短期大学

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

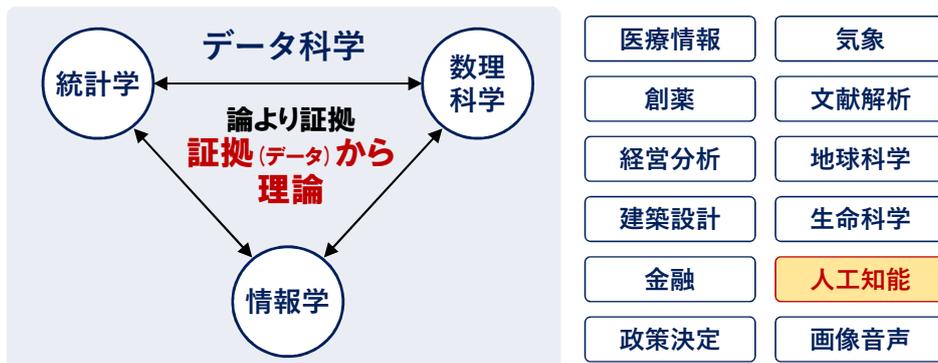
このような政策を実施すべく、文部科学省は数理・データサイエンス・教育強化拠点コンソーシアムという組織を構成しました。現在200を超える大学が参画しています。

そして京都大学は、このコンソーシアムの近畿ブロック拠点校になっていて、データサイエンス教育のリーダーシップをとり、主に近畿圏内の大学と一緒にデータサイエンス教育を展開し、普及させています。

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、
数理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

Twitter公式アカウント [@cireds2020]

5

京都大学全学共通教育ではデータサイエンスのことをデータ科学とよんでいます。データ科学は新しい学問ですので、その定義がまだ明確に定まっているわけではありません。

京都大学全学共通教育ではデータ科学を、データを収集してコンピュータで管理し、数理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、ととらえています。

つまり、データ科学は統計学・情報学・数学の融合分野ととらえています。

論より証拠ということばがありますが、このことばをもう少し深く考えると、証拠であるデータから筋道を立てて理論を導く、ということになります。

データ科学は現代における「論より証拠」のための学問といえます。

データ科学は、ここにあげるような広い分野の学術で基礎となる学問です。

そして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。

統計学は数学の一部ではありません

日本において令和3年度以前の高校までの学習指導要領では、**統計的手法は数学の一部として組み込まれています**

- 高校では**数学I「データの分析」**、**数学B「確率分布と統計的推測」**
- 歴史的には、著名な数学者が統計的手法の構築に大きく貢献したことは事実です

しかし

統計学は数学を基盤とする別の学問です

- 「日本統計学会」と「日本数学会」は別組織です

さらにITの発達により、**情報学**との関係が密接になり、

現代は **統計学** × **情報学** × **数学** = **データ科学** のどれもが重要です

6

ところで令和3年度以前の高校までの学習指導要領では、**統計的手法は数学の一部として組み込まれています**

数学Iの「データの分析」、数学Bの「確率分布と統計的推測」の単元が統計手法にあたります

歴史的には、著名な数学者が統計的手法の構築に大きく貢献したことは事実です

しかし、**統計学は数学を基盤とする別の学問です**。「日本統計学会」と「日本数学会」は別組織です

さらにITの発達により、**情報学**との関係が密接になり、

現代は、**統計学**、**情報学**、**数学**、そしてこれらが合わさった**データ科学**のどれもが重要です。

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)¹⁾



Johannes Kepler
(1571-1630)²⁾

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?”

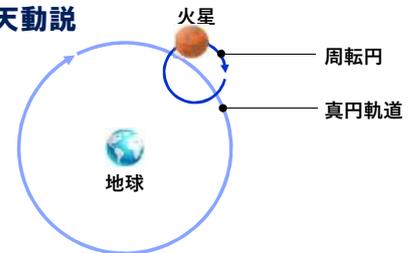
地動説の下で、火星の軌道が楕円だと仮定すると
データが説明できることが分かった!

ケプラーの法則（第1法則）

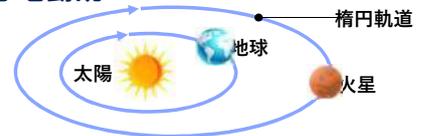
惑星は軌道は楕円である!

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

データ科学は多くの学問の基礎であるといいましたが、自然科学もその一つです。皆さんの中には、物理学の法則を単なる公式とと思っている人がいるかもしれませんが、しかし、数学の公式と違って

物理学の法則は、古来多くの人たちが自然現象を観測して得たデータから導き出された結果なのです。

たとえば、惑星の軌道は太陽を中心とする円なのではなくて、楕円になっているのだと高校時代に学習したと思います。

惑星の軌道が楕円になっていることを発見したのは、ヨハネス・ケプラーでした。ケプラーの発見には、その前の時代の天文学者であるティコ・ブラーエが火星の運行を正確に記録した、当時としては高精度なデータがあったからだといわれています。そのデータを見たケプラーは、火星は楕円の軌道を描いて動いているのではないかと着想します。

このケプラーの法則は後にニュートンの万有引力の法則へと発展します。

そして現代の高度な科学をさらに発展させるには、膨大なデータを高速にかつ適切に処理することが必要不可欠です。

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來 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%		



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

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

データを集めることの重要性は、医療現場や政策決定にとっても極めて重要です。19世紀にロンドンにおいて特定の井戸周辺にコレラが多発する事件が発生しました。

イギリス人の医師で、疫学の創始者と呼ばれているジョン・スノウはロンドンの街のコレラの発生地点のデータを集め、それを地図の上に表すことで特定の井戸の周辺にコレラが多発していることに気づきました。

そしてその事から同じ地域でも水道供給会社によってコレラの発生率に大きな違いがあることを発見し、コレラの感染症対策が大きく進みました。

これで皆さんはお気づきでしょう。最近の世界各国のウィルス感染政策においても、データの収集によって、感染経路を特定し、感染拡大の防止につなげようとしています。

このとき、たとえば検査に対する偽陰性の割合がどれくらいかを見積もるには、適切なデータを基礎にして、条件付き確率をはじめとする数学を利用することになります。

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



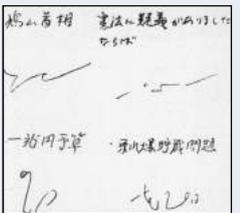
衆議院本会議
*5



新システム



*6



*7

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

統計学

数学

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

音声記録と議事録の違いを数式化

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

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

特に最近では、データ科学と人工知能AIとの密接な関連が注目されています。昨今では、人間の話す言葉を認識する技術は身近になりましたが、京都大学では10年前にすでにその技術を実用化し、衆議院の会議録作成システムに応用しています。

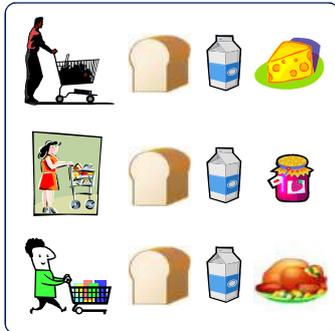
会議録はこのような速記文字という特殊な文字で作成され、そのあとに通常の文字に書き直されます。

衆議院の会議録作成システムでは、人間が話す言葉の音声データから直接、普通の文字で記録されます。

この技術の背景にあるのは、膨大な量の過去の音声データと議事録データです。そして統計学、情報学、そして数学を融合することでシステムは設計されています。まさに本学のデータ科学そのものということになります。

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？

10

データ科学は今やビジネスの世界でも必須です。例えばスーパーマーケットやデパートで買い物をして、レジで支払いをするということはそのお店に購買データを渡していることにもなります。お店の方は、膨大な購買データを分析し、よく売れている商品、あまり売れなくなってきた商品、特定の季節によく売れる商品、一緒に売れることが多い商品などを知ることにより、経営を効率化しようとしています。これは、売れ残りを回避するなどエコロジーにもつながります。

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

データサイエンススクール (課外)

11

全学共通教育でのデータ科学分野、情報群の科目の一例をお見せしておきます。統計と人工知能といった先進的な内容も提供しています。もっと詳しい情報は、履修案内や、データ科学センターのパンフレットをご覧ください。データ科学センターのパンフレットは国際高等教育院や各学部の教務事務室においてあります。

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準拠

■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
- 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
- 統計解析ソフトを用いたデータ分析を自習形式で習得する



13

データ科学分野の中心となる科目は、「統計入門です」全学部・全回生向けの基礎的な科目です。

この科目は文科省の数理・データサイエンス・AIモデルカリキュラム(リテラシーレベル)に準拠しており、合格すると特別に修了証を受領することができます。

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しています。

統計と人工知能, データ分析基礎, データ分析演習 I, II

■ 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に準拠

■ 4科目中1科目以上の合格者は**修了証**を受領可能

統計と人工知能

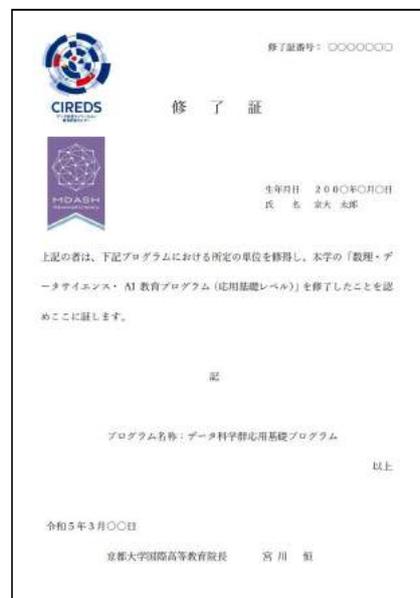
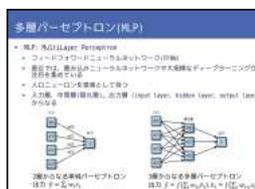
- 「統計入門」の発展として, より高度な内容を習得
- **人工知能・機械学習**の基本的な考え方や手法を習得

データ分析基礎

- 「統計入門」より先の分析法や推測アルゴリズムを習得
- 分析の使いどころを**統計解析ソフト**を用いて**実践的に**習得

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
- 今後専門とする分野に合わせて**幅広い選択**ができるように, 異なる内容の**科目を揃える**



基礎的な内容を習得した後に, さらに進んだ内容を提供する科目として, 統計と人工知能, データ分析基礎, データ分析演習I, IIも用意しています。

これらの科目は文科省の**数理・データサイエンス・AIモデルカリキュラム (応用基礎レベル)**に準拠しており, **1科目以上の合格者は修了証を受領することができます。**

内容は, 人工知能や機械学習の基本的な考え方を習得したり, 大規模なデータの分析を統計ソフトを用いて実践的に習得するなど, より高度で**実践的なもの**になっています。

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時が来ます。

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/>

Twitter公式アカウント [[@cireds2020](#)]

14

いま説明しましたように、データ科学は様々な学問や技術の基礎となっている、まさに読み書きそろばんに匹敵する基本的な学問であり、誰でも将来必要となる時が来ます。

皆さんのお手元にある案内にありますように、データ科学イノベーション教育センターでは主に統計学とその発展科目を提供しています。

皆さんの所属する学部・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。

ただし、全学共通教育で学習しなければならない内容は、学部・学科によって異なることには注意してください。

ここにありますアドレスもしくはTwitterの公式アカウントでも情報を載せています。データサイエンススクールなど講義以外のイベントの情報も発信していますので、参考にしてください。

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。

データ科学イノベーション教育研究センター (CIREDIS) 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：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・スクール
- データサイエンス入門講座

詳しくは下のHP または Twitter公式アカウント [@cireds2020] でご案内
<https://ds.k.kyoto-u.ac.jp/category/event/>

15

繰り返しになりますが、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。

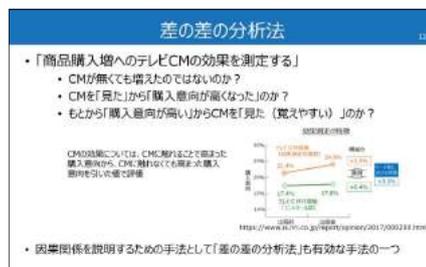
そのため単位数や時間割などの制約によって、データ科学に関する科目を希望通りに履修できない可能性もあります

そのような場合のために、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

MOOC「統計の入門」

全国向けe-learningサイトJMOCで毎年度公開

<https://www.jmooc.jp/>



- 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
- 最後まで履修すると**修了証**が発行されます

16

まず、統計学の基礎については、全学共通科目「統計入門」の要点だけを学習できるようなe-learning教材として「統計の入門」を提供しています
毎年度、全国向けのJMOCというサイトでご覧いただけるようにしています。
「統計入門」の予習に利用することもできます
また、最後まで履修すると**履修証明書**が発行されます

Pythonプログラミング自習キット

https://kubar.rd.iimc.kyoto-u.ac.jp/intro_python/

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 Collaboratory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



プログラミングについては、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をKubarというサイトで配布しています。
Internet環境とWebブラウザだけ用いて自分のパソコンで自習できるようになって
います。

データサイエンススクール(随時開催)



12/3(土) 10:00-17:00 12/10(土) 13:00-17:00
 12/17(土) 13:00-17:00 12/24(土) 13:00-17:00
 講師: 速川 徹 (アクトクエストテクノロジー株式会社)

オンライン開催
 (Zoom)
 参加費無料
 定員: 各50名

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「数値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新動向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出来る機会を提供するよう企画いたしました。

12/3(土) 10:00-17:00
 スクール7B: プログラミングのためのPC操作とPython入門
 対象レベル: PCの基本操作はできる
 プログラミングに興味がある

12/10(土) 13:00-17:00
 スクール7C: 機械学習トレーニング
 対象レベル: Pythonの関数定義を用意できる
 Pythonの基本文法を理解している

12/17(土) 13:00-17:00
 スクール8D: 機械学習トレーニング
 Webアプリ開発編
 対象レベル: 前編までの内容を理解している
 文字認識の基礎を理解している
 サンプルコードを実行できる

12/24(土) 13:00-17:00
 スクール8E: 機械学習トレーニング
 Semantic Segmentation 編
 対象レベル: 前編までの内容を理解している
 ニューラルネットワークの基礎的な知識がある

12/17 スクール8Dの終了後にIT業界訪問を行います。IT業界に興味がある方はぜひ！

募集対象: データ科学実践IV対象スクール
 対象: 京都大学に所属する学生・教職員
 優先対象: 「データ科学実践IV」履修生

お申し込み方法:
 専用フォームからお申し込みください。
 順次情報をお知らせいたします。

<https://ds.k.kyoto-u.ac.jp/2022winter/>



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

また、開講科目だけではカバーできない実践的な内容、先進的な内容を学びたい人には、データサイエンススクールという課外学習を開催しています。

春期休業、夏期休業、週末などを利用し、学内の研究科・研究所・附属病院、あるいは他大学や企業から先生方をお招きして、1日~4日間の集中講義を行います。

詳細はデータ科学センターホームページで確認してください。

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の
ホームページで確認してください。

19

講義が受講できない人や、折角データ科学を勉強したから資格を取っておきたいという人のために

京大オリジナルという会社にデータサイエンス入門講座を開講していただいています。

「統計検定」に合格するための講座ですので、有料ですが、一般財団法人統計質保証推進協会との連携セミナーになっています。

詳細は京大オリジナルのホームページ、データ科学センターのホームページ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

Twitter公式アカウント [@cireds2020] で情報発信中

以上、全学共通教育でのデータ科学について説明いたしました。

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

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するのは間違いです

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学習をサポートします

全学共通教育の科目を担当するほか

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ので活用してください

どうか積極的に現代社会の素養であるデータ科学を学んでください。

図版等出典(最終確認日:2023年2月14日)

- *1 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
 - *4 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
-

21

https://upload.wikimedia.org/wikipedia/commons/9/9f/Shugiin-system_Shorthand_01.jpg