

全学共通教育

ビッグデータの時代
データを科学を学ぼう

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

お手元に届いていると思います

その裏側には、データサイエンス。

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

あの宅配の裏側にも

最高の買い物物の裏側に

健康維持の裏側に、

ともだちの始まりの裏側に、

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

教育改革に向けた主な取り組み

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

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、高校等での活用促進

リテラシー

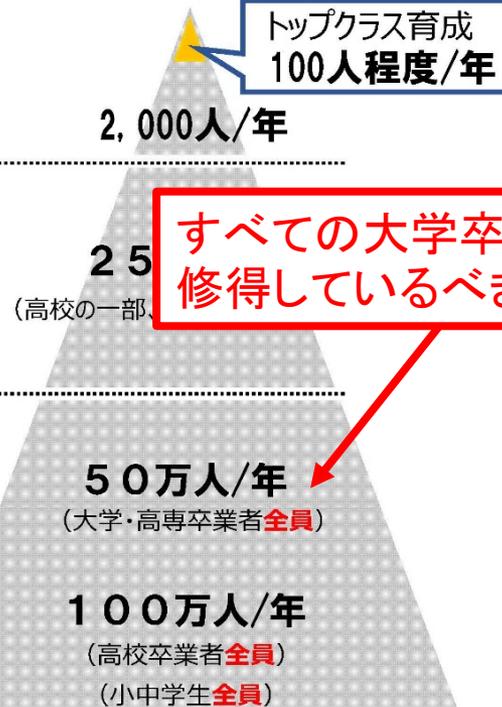
学習内容の強化

- 大学の標準カリキュラムの開発と展開（MOOC※活用等）
- 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）

育成目標【2025年】

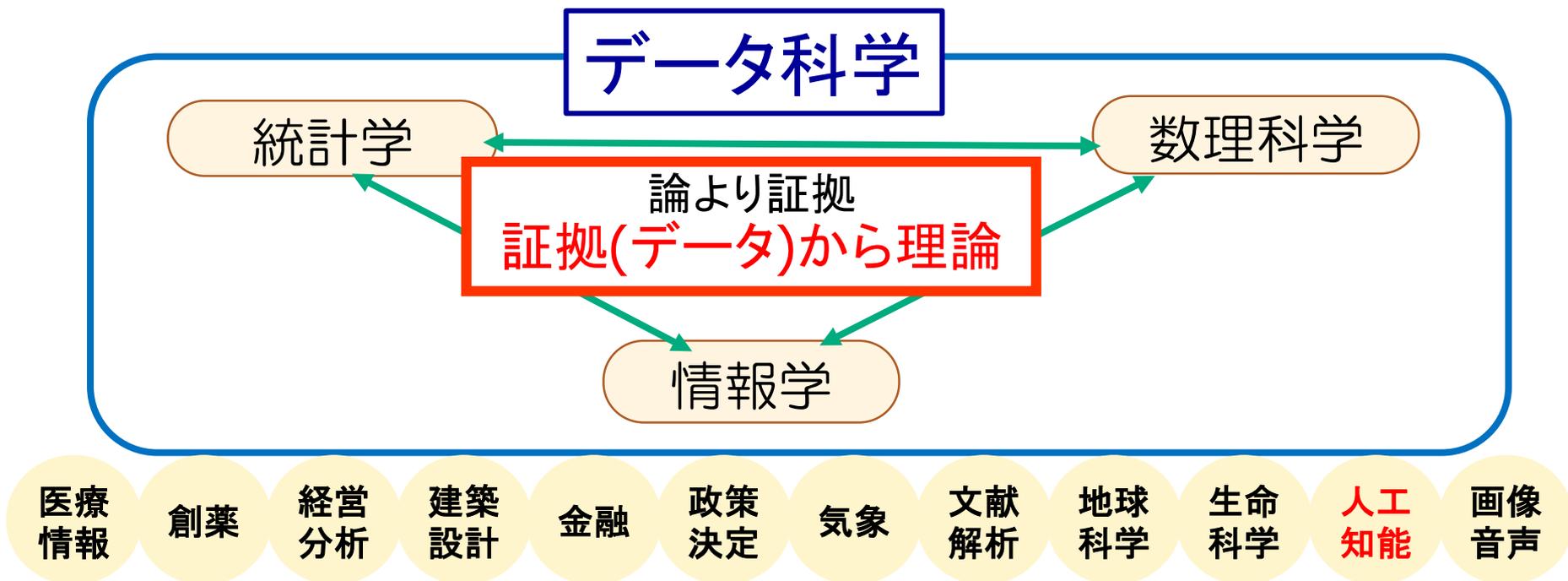


すべての大学卒業生が
修得しているべき素養

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

統計学は数学の一部ではありません

- 日本における高校までのカリキュラムでは、**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
- 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

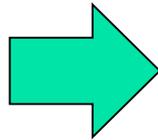
データ科学と人工知能(AI)

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

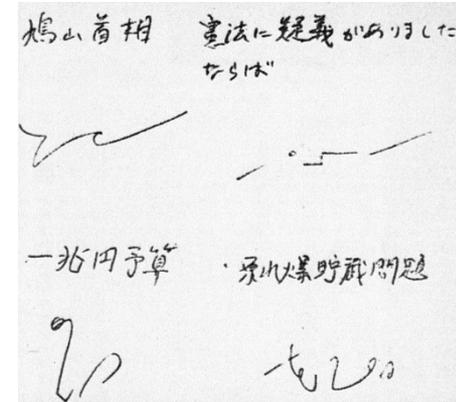


*3

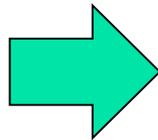
従来



*4



*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

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

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

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

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!

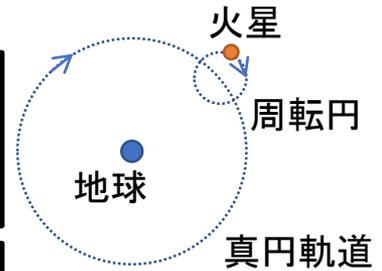
ケプラーの法則(第1法則)

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

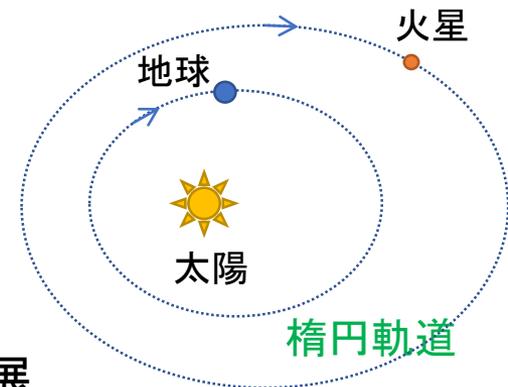
地動説の場合は、真円軌道だと誤差が大きい



Tycho Brahe(1546-1601)



*2



*1

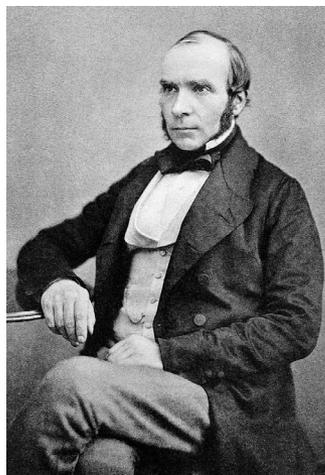
データ科学と医療・政策

1854年ロンドンにおいて
特定の井戸周辺にコレラ
多発

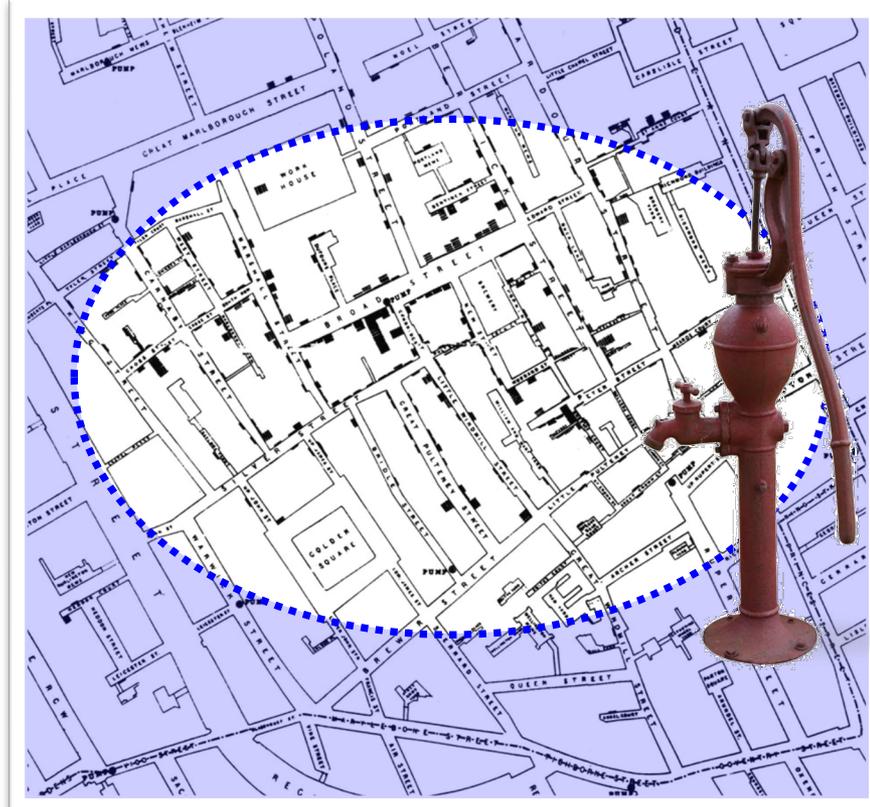
データを集めることにより
同一の区域でも、水道供
給会社によってコレラ発
生率に大きな違いがある
ことを発見



コレラ対策という政策が進む



John Snow
(1813-1858)*6



*7*8

流行ピーク、8割がインフルエンザの外来

検査	インフルエンザ		計	
	あり	なし		
陽性	1440	8	1448	陽性的中率 99.4%
陰性	160	392	552	陰性的中率 71.0%
計	1600	400	2000	
	感度 90%	特異度 98%		

偽陽性

偽陰性

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/>



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 (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">• 数理・データ科学のための数学入門 I・II• 統計入門• 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)
専門基礎レベル (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">• 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)

図版等出典(最終確認日:2020年3月25日)

- *1 https://commons.wikimedia.org/wiki/File:Tycho_Brahe.JPG
- *2 <https://commons.wikimedia.org/wiki/File:Kepler.png>
- *3 [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)
- *4 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>)
- *5 衆議院式速記の一例
https://commons.wikimedia.org/wiki/File:Shugiin-system_Shorthand_01.jpg
- *6 https://commons.wikimedia.org/wiki/File:John_Snow.jpg
- *7 http://matrix.msu.edu/~johnsnow/images/online_companion/chapter_images/fig12-5.jpg
- *8 Adobe stockより購入 (AdobeStock_5473397.jpeg)