

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
自己点検・評価報告書
ーリテラシーレベル教育編ー

令和3年5月

はじめに

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（以下「センター」という）は、平成 28 年に文部科学省による「数理及びデータサイエンスに係る教育強化」事業において拠点校の一つとして京都大学が選定され、共通政策課題（数理及びデータサイエンスに係る教育強化）として予算措置されたことを受け、翌年に国際高等教育院附属センターとして設置された。

いわゆる第 4 次産業革命の前触れとも言える AI・ビッグデータ解析などの新技術の急速な進歩は、学術全般・産業界のみならず日常生活の至る所に大きな変化をもたらしており、これらの技術革新の根幹である情報学・統計学・数理科学に対する基本的な理解は、文理を問わず、社会を支える広範な人材にとっての基礎的な教養となりつつある。

京都大学には、情報・統計・数理に関しトップレベルの研究業績を有する教員が多数在籍し、その学識等を基にそれぞれの学部等で教育を行っていた。センター設置の目的は、これらの教員が連携し、論理力の涵養を根幹とした 21 世紀の基礎教養としての情報学・統計学・数理科学に関する基盤教育を、学部における全学共通教育から大学院における高度専門教育までの各段階に応じ、効果的かつ全学的に提供できる体制を整備し、データ科学者の養成や産業全般で同時進行する第 4 次産業革命をトップレベルで支える人材育成を行うことにある。具体的には、センター設置前から開講されてきた「統計入門」や「数理統計」に加え、専門教育基礎教育、専門教育、高度専門教育に関し、学部・大学院のニーズを聴取しつつ、開設科目を具体化し、カリキュラム整備・科目の設計等を行うとともに、特に多数の受講者が見込まれる科目について e-learning 教材を開発し、体系的な情報学・統計学・数理科学教育を展開する計画である。また、拠点校 6 校による「拠点大学コンソーシアム」を形成し、連携体制を構築していくなかで、積極的な情報提供や役割分担などを通じ、全国的なモデルとなる標準カリキュラムの作成にも貢献する。

本報告書では、4 年間のセンター活動のうち、「統計入門」を中心としたリテラシーレベルの教育について、センター設置前の科目の設計段階を含めて報告する。

令和 3 年 5 月

1. 京都大学における学部学系向け全学共通科目

京都大学の学士課程は

総合人間学部、文学部、教育学部、法学部、経済学部、
理学部、医学部、薬学部、工学部、農学部

の 10 学部から構成されている。通則第 16 条では、学部の教育課程において、開講対象による区分として全学共通科目及び学部科目とし、教育目的・内容による区分として教養科目及び専門科目としている。また、同第 18 条では、学士課程の科目、授業、修業年限及び在学年限は、各学部が定めることになっている。そして、各学部の学士課程の編成方針に基づき、教養科目及び外国語や専門基礎科目といった各学士課程に共通する科目の企画及び実施を担当する組織として国際高等教育院が設置されている。センターはその目的から、国際高等教育院の附属センターの一つとして設置された。

全学共通科目は異なる教育目的をもつ下の 8 つの科目群に区分されている。

人文・社会科学科目群、自然科学科目群、外国語科目群、情報科学科目群、
健康・スポーツ科目群、キャリア形成科目群、統合科学科目群、
少人数教育科目群

さらに各科目群は分野の科目群に細分されている。例えば、データ科学が属する自然科学科目群は

数学、データ科学、物理学、化学、生物学、地球科学、図学
に細分される。各科目群あるいは分野の科目を企画し実施する組織として、各学部から選出された委員会からなる「部会」が設置されている。データ科学分野については、センターが本格的に活動を開始する平成 29 年度までは分野としては独立していなかった。後述するように、「統計入門」を開講するにあたり、当初は数学部会と情報系部会、その後生物部会が加わって合同小委員会を設置して科目が設計されたので、「統計入門」はこれらの部会が所掌する科目群のどれかに置くことになるが、開講後 3 年間は「数理統計」が数学科目の一部であったこともあり、数学科目としての開講であった。平成 29 年度からは「統計入門」を含むデータ科学科目は、自然科学科目群の一分野となった。全学共通科目で履修しなければならない単位数については、どの学部も自然科学群科目の方が情報科目群より多いため、新規参入のデータ科学科目が自然科学群科目に入ったことは、学生にとって履修しやすい状況を生み出したといえることができる。

2. リテラシーレベル教育「統計入門」の開講と変遷

本学の学部学生向け全学共通教育におけるデータサイエンス教育の中心となる科目は「統計入門」である。本章では同科目の開講までの経緯を、参考文献[1][2]の該当箇所を抜粋し、本報告書の目的に沿って編集・修正したものを掲載したのち、開講後の変遷について説明する。

2.1 開講までの経緯

全学共通科目「統計入門」が開講された直接の経緯は、国際高等教育院が設置に先立って平成 25 年度に設置された企画評価専門委員会基礎教育検討ワーキング・グループにおいて、分野別部会の検討を始めるにあたって、医学部から統計学についての要望が寄せられたことに始まる。確率論をベースに数学的な定理と証明付きの議論を展開するのでなく、数学嫌いにならないルートで、使える統計あるいは騙されない統計を身につけることができる科目を提供して欲しいという趣旨であった。これを受けて検討が行われた結果、基礎教育検討ワーキング・グループの報告では、「Ⅲ. 基礎教育の改善のために取り組むべき分野別の課題」の「1. 科目編成の基本的な考え方等」において、一般論として

国際高等教育院において提供される基礎教育科目と、各学部において提供される基礎的な専門科目の区分は、当該科目の内容が、自然・応用科学研究を行う者にとって共通の基盤となるものか、一定の専門教育にとって基礎となるものかによって行うのが原則であると考えられる。ただ、基礎科目の履修においても、専門科目との関係が明確である方が、学生の学修意欲を高めることができると考えられることから、クラス指定科目においては、当該クラスの学生の専門分野とのつながりを意識した授業を行うなどの工夫を行うことが望ましい。

との記述がなされ、また、「2. 各分野の検討」の「(1) 数学分野②検討すべき課題(科目内容)」において、特に

「数理統計」は、基礎科目において数学的な基礎を中心に教育がなされている一方で、応用としての統計学については各専門の特殊性に対応するために、各学部の専門科目として開講されている場合が多い。ただ、前者においても、数学的な基礎に応用的な側面を一定程度加えていく改善が考えられるが、そのためには、各学部の協力が必要である。

と述べられている。これを受けて、**数学部会**において、平成 27 年度以降の数学分野のモデル科目編成案作成の一環として、統計科目の見直しが必要であるとの認識のもと、統計小委員会を設置し検討を始めることが決まった。その後、**情報系部会**でも、統計ソフトを扱う際に知っておくべき統計の基礎知識が話題となり、両部会の合同小委員

会として開催された。さらに**生物部会**においても、理学部の生物系教員から、「数理統計」とは違う内容の統計科目の必要性が提起され、第3回小委員会からは**数学部会・情報系部会・生物学部会の合同小委員会**として議論することになった。

全学共通科目の自然・応用科学系科目群において開講されている統計科目は主に以下の3つである。

「確率論基礎」と「数理統計」は、それぞれ2回生前期と後期に理学部2クラス、工学部4クラス、農学部1クラス指定されていて、農学部の「数理統計」がシラバスも異なる内容で農学部の教員が担当している以外は、共通のシラバスで、主として専任教員または非常勤講師の数学系教員が担当している。工学部の「数理統計」1コマは、情報学研究科の教員の担当である。医学部のクラス指定科目に「数理統計」があり数学系教員が担当しているが、これは対応して「確率論基礎」がクラス指定されておらず、内容が異なる。この他に、平成25年度からの新規科目として「文系のための数理統計入門」が前後期1コマずつ数学系教員によって提供されている。

学部の教員が教えるのではなく数学系教員が教える場合は、それぞれの専門分野での応用統計を意識して教えられているわけではない。そういった点よりも、統計的推論に確率論に基づく数学的基礎を与えることを、通常の統計学の教科書よりさらに徹底して教えることに力点が置かれる。これは、統計の基礎となる数学が、1回生時に学ぶ微分積分学のよい応用の舞台であるという点から、本学の数学教育の見識の高さを示すものではあるが、必ずしもすべての学部で、そこに力点を置く統計科目が求められているわけではない。

一方、統計の基礎付けへの数学の応用ではなく、専門分野の諸科学への統計の応用という点を考えるとき、数学者には、統計を自分の研究に応用するという経験がなく、このことが、数学者が応用の実例を教えることを困難にしている。

統計に対するニーズの強い学部学科は、全学共通科目の「確率論基礎」「数理統計」をクラス指定するのではなく、自前の学部科目を持っている場合が多い。

経済学部では、1回生前期に数学科目を学ぶのと並行して「基礎統計学」を学習するカリキュラムになっている。この科目は2回生以上の後期科目である「経済統計学」「計量統計学」に接続するように考えられているので、1、2回生で200名を超える履修登録がある。経済学部としては、全学共通科目にニーズがあるわけでもなく、自分の学部だけで手一杯であり、全学共通科目に統計科目を提供する余裕はない。

農学部では、6学科のうち2学科で「生物統計学」が、1学科で「農林統計学」がどちらも2回生後期科目にある。これらには、演習（実習）のコマが並行して開講されている。後者の学科では「数理統計」の履修者も多い。残りの3学科のうち、1学科は「確率論基礎」「数理統計」のセットの履修であるが、2学科では統計科目の履修は少ない。

工学部では、2 学科が「確率論基礎」「数理統計」をクラス指定せず、自前の講義を持っている。残りの 4 学科はクラス指定であるが、新しい科目に対する興味が示された。そのうちの 1 学科は、さらに専門の統計科目があるため、重複しているのではないかという議論がでている。

教育学部では統計を使う分野では専門で教えているが、教養・共通教育の中に基礎的な部分を教える科目があれば歓迎する。

医学部では、医学科からは、統計学の基本的な考え方を、数学的基礎付けよりも応用例に重点を置いて教えてほしいという要望が、一方、人間健康科学科からは、数学的基礎を平易に教えてほしいという要望が出た。どちらの学科も、「数理統計」までの数学的内容を求めているわけではなく、また、医療統計や疫学統計といった専門的内容を全学共通科目に求めているわけでもないという点で共通している。

薬学部も、「確率論基礎」「数理統計」ではないものを求めているという点は医学部と同じであるが、これまで、専任教員が学部科目として「バイオサイエンス統計基礎」を講義していた時期があった。現在は非常勤講師の担当であり、平成 27 年度からは非常勤講師が続けられないため、全学共通科目での適切な科目の提供を期待している。

理学部では生物系から「確率論基礎」「数理統計」ではない科目設計を求める声が出ている。

これ以外の、主として文系の学部の場合は、文学部や教育学部のように、統計を必要とする専門分野を持つところがあるが、全学共通科目としての統計に強く期待する声は聞こえてきてはいない。しかしながら、「文系の数理統計入門」には前後期ともに 70 人程度の履修者があるので、文系の学生にとっても統計のニーズはあると考えるべきである。

以上述べてきた科目提供の現状をもとに議論を行った結果、以下の基本方針が承認された。

- ・ 平成 27 年度より、「数理統計（医学部）」は廃止、理学部のクラス指定の「数理統計」（現行 2 クラス）は 1 クラスのみとする。「文系のための数理統計入門」（現行前後期 1 クラスずつ）は廃止する。代わりに「統計入門（仮称）」（半期 1 コマ）を 4 クラス用意する。
- ・ 「統計入門（仮称）」は医学部医学科、医学部人間健康科学科、薬学部、理学部等でのクラス指定を検討する。同時に、クラス指定以外の履修者も想定し、前後期にバランスよく配置するものとする。
- ・ 「統計入門（仮称）」の科目内容は、「数理統計」とは差別化する。「数理統計」は 1 回生の数学科目および 2 回生前期の確率論基礎の履修を前提とする 2 回生後期科目であり、統計学の基本的事項を数学的基礎付けの理解と共に学ぶことを目標と

する。「統計入門（仮称）」は、統計の考え方を知ることと、統計の応用の実例について学ぶことを主眼とし、数学的な基礎付けについては、対象学生のニーズや予備知識に合わせて適宜取り入れるものとする。

さらに具体的な**実施方針**として以下を設定した。

- (1) 全学共通科目でのニーズのある学部代表者若干名を加えた科目設計・教材開発ワーキング・グループを構成し、平成 27 年度からの授業実施に向けて科目設計・教材開発を行う。
- (2) 単なるシラバスの策定ではなく、OCW や e-learning も含めた教材づくりを行う。教科書を作成することが目的ではなく、各回の授業がどのような組み立てになるかを授業担当者がはっきりとイメージできるための手引き（教科教育法での指導案に相当するもの）をつくる。
- (3) 統計がどのように使われるのかという実際面からの導入を重視し、専門へのつながりの多様性に配慮した教材づくりを行う。それによって、クラスごとの学生が自分の興味分野との関連を意識しつつ学習できることを目指す。
- (4) 科目目的（学生が何を学び、何が出来るようになるか）を明確化し、それに基づいた成績評価の基準と方法づくりも行う。

「確率論基礎」、「数理統計」、「統計入門（仮称）」を合わせた統計科目全体を統括する統計教育特別部会を国際高等教育院企画評価専門委員会に設置し、科目設計・教材開発チームをその下に置く。統計教育特別部会は、平成 26 年度中の作業においてチームを支えると共に、平成 27 年度からの授業実施の責任を持つ。さらに、平成 27 年度前期に、授業の検証を行い授業内容の必要な修正を行い後期授業に活かすとともに、平成 28 年度以降の統計科目の実施体制を検討する。地球環境学堂の吉野章准教授が平成 25 年度に新しく 1 年限りで提供した「社会統計学」に 300 名近い履修者がいることがわかり、科目設計の中心となっていただくことをお願いし、快諾とともに科目設計 WG の座長を引き受けていただいた。

2.2 情報学研究科教員による開講

「統計入門（仮称）」の担当は、情報学研究科知能情報学専攻の教員が 6 クラスを提供することに合意され、平成 27 年度開講を目指して準備を始めることになった。情報学研究科教員が担当することになったのは、同研究科設立時に総合人間学部から生物系教員ポストが移される際に全学共通科目提供も同時移されることになったことに由来する。このため、もともとは生物系科目を提供していたが、生物系教授が定年退職するにあたり、同研究科の専門領域により近い科目を提供するように同研究科と国際高等教育院の間で交渉が続けられていたところに「統計入門」開講が重なり、この

科目を同研究科の教員が担当することで合意したものである。

情報学研究科で「統計入門」を担当する予定の教員からなるワーキング・グループが構成され、講義資料作成に着手した。「社会統計学」の講義資料(スライド)をベースとしながらも、受講する学士が全学部に通じることを想定し、高等学校の「数学」の内容も参照しながら、国際高等教育院が開講するリテラシー科目の役割である

- ・ 学生が高等学校の教育から大学での専門基礎教育へ円滑に移行するための橋渡し
- ・ 学生が自ら知的世界を広げようとするとき、その介添としての役割

という 2 点もあわせて検討が行われた。具体的には、標準的な統計学の内容である推定と検定に加えて、医療分野や情報学では必須の 2 元分割表、実験を行う学問では必須である検定、予測を必要とする分野で用いられる回帰から構成されることし、最低限必要な数学・確率論の内容を交えながら講述する内容とした。

小委員会では、「統計入門」は座学ではあるものの、前節に述べた実施方針(3)を踏まえて PC を用いてデータの統計処理をするという演習を組み込むことが決められた。ワーキング・グループでは、この演習に利用する PC は本学の BYOD 化の方針を念頭に置いて、各自の所有する PC あるいは学術情報メディアセンターにある端末を用いることとし、座学の時間を確保するために、講義時間中に演習の内容を説明した後、自宅学習の一部として演習を設定することとした。演習には統計処理におけるプログラミングの要素が全くなく、メニュー選択方式で操作が可能な JMP を採用し、医学研究科・薬学研究科・国際高等教育院の 3 部局が共同で全学ライセンスを購入することで対応した。JMP は医薬関係で統計データ処理に際して認定を受けていることもあり、「統計入門」以外の教育にも利用できる可能性も含めて採用された。

平成 29 年度までの 3 年間は、**各年度の履修者数は 800 名程度**であった。当初は振れ幅の小さいシラバスに基づいた、統一された講義形式の実施を行っていた。一方で、サイズが 200 名近くになるクラスもあり、さらに 1 クラスの中には文系・理系・医学系などが混然として登録・受講するために、**数学を中心とした受講生の基礎学力のばらつきが目立ち**、説明の方法や速さなどについてベースラインが定まらず、授業の進行や適切な評価という点で課題を生じることになった。

2.3 センターの設置と科目見直し、充実化

平成 29 年に共通政策課題「数理及びデータサイエンスに係る教育強化」が予算措置され、国際高等教育院に附属データ科学イノベーション教育研究センターが設置された。この予算に加えて京都大学独自予算(IR 経費)を合わせて、特定教員 6 名が雇用可能となった。さらに情報学研究科知能情報学専攻の担当講義見直しにより、データ科学群科目の担当者を新たに 2 名増やすこととなり、平成 30 年度より合計 14 クラスを開講する体制が整った。センターには**教育学・理学・医学・工学・農学・情報学**を専

門とする 6 名の特定教員が次々と着任し、自分の専門分野に合わせて研究科を兼務することで大学院教育にも寄与することとなった。

このような体制の充実により、まず 1 クラスの人数を減らして大規模クラスの持つ弊害を除くことが可能となった。さらに受講者数の増加により**数学を中心とした受講生の基礎学力のばらつきが拡大する**ことが予想されることから、実施方針(3)について、学部・学科のクラス指定制度と担当者の専門分野の違いなどを利用して、クラスごとの授業内容に変化を持たせることを実現することとした。特に、数学 III の内容を課さない入試（いわゆる文系入試）を経て入学した学生向けのクラスを前期・後期にそれぞれ 1 クラスずつ開講することとした。規模の拡大と専門性の拡大とともに、自由度を確保する必要性も生じたため、「**授業計画と内容**」を 14 区分から 9 区分に変更したうえで、**順番も明示しないよう変更**することになった。さらに、PC を用いた演習的内容についても、JMP の全学ライセンスはセンターが購入した上で利用し、さらにはプログラミング言語 R を用いるクラスを 1 クラス設定し、多様な学生のニーズに対応するようにした。

一方で、「統計入門」の内容が多すぎるのではないかと、との意見が担当教員から出たことを受け、「統計入門」の内容から平均差の検定、分散分析、回帰と因果推論を削除し、それらに加えて新たな内容を追加した「**続・統計入門**」を新規に開講することとした。「続・統計入門」は前期と後期に 1 クラスずつ開講することとした。

平成 30 年度、平成 31(令和元)年度ともに履修登録者の数は 2 クラス合わせて数十名程度であった。「統計入門」では飽き足らない学生に対しその次の科目を提供する、という役割は果たせているが、リテラシーとして多くの学生が受講したとは言いにくい状態であった。履修登録者数がそれほど多くなかった理由として、統計入門に関しては複数の学部において履修を推奨されているのに対し、科目数の関係などから**続・統計入門**までを推奨しにくいことが一因と考えられる。ちょうど令和元年に「**AI 戦略 2019**」が内閣府・文部科学省から発表されたことを受け、令和 2 年からは学生自体が強く興味を惹かれるであろう人工知能の内容を增強し「**統計と人工知能**」という名前に改称し、引き続き前期と後期に 1 クラスずつ開講した。

リテラシー科目としては、文系入試を経た学生が「統計入門」などのデータ科学群科目を履修する際に必要な数学的知識を補完するための科目として、「数理・データサイエンスのための数学入門 I・II」を設計し、開講することとした。このような体制の充実に加えて、全学向けの広報などの成果もあり、**令和 2 年度実績では約 1,200 名弱まで履修者数を増加**させることに成功した。「統計入門」担当者は頻繁に FD(Faculty Development)を行い、互いに助け合いながらも質の高い講義を提供することができた。

2.4 COVID-19 対応を経て飛躍したオンライン・オンデマンド教材

令和 2 年度前期講義は、COVID-19 の蔓延を受けて波乱の幕開けを迎えることとなった。国際高等教育院では 5 月の連休明けまで休講決定し、その間は各講義に沿った自学自習の方向付けをすることになった。このような感染対応策は、一般には学生の学修時間と内容を確保するという問題を生じたが、「統計入門」については、後述するオンデマンド教材である「統計の入門」などを最大限に活用することができたため、比較的スムーズな対応を実現できた。国際高等教育院では 5 月の連休以降も、科目担当者がオンライン型・オンデマンド型を選択しながら遠隔講義として開講することになったが、いずれのアプローチであっても**録画された動画コンテンツが蓄積され、さらなる改善プロセスへの重要な基礎資料**ともなった。特に、オンデマンド型遠隔講義では、企業内研修で先行してきた、最大でも 15 分程度単位の動画に区切られた動画コンテンツ教材を主体とするマイクロラーニング (Wang-Audia & Tauber, 2014, Meet the Modern Learner: Engaging the Overwhelmed, Distracted, and Impatient Employee, Bersin by Deloitte) の手法も活用され、従来から指摘されてきた細切れ時間の有効活用にもつなげることができた。さらに、デジタルネイティブである学生からは「内容によって、高倍速閲覧・低倍速閲覧・一時停止・繰返し受講ができる」「移動時間も含めて自身の都合に応じて受講」と前向きに捉えられる意見が寄せられた。こうした結果として、評価の主対象となった最終レポートの質が飛躍的に向上し、オンライン型・オンデマンド型いずれであっても遠隔講義の実施に自信を持つことができ、今後の活用も大いに期待できることとなった。実際、令和 3 年度前期講義は開講当初は対面形式での実施であったが、教室に入る学生数を教室の収容人数の 2/3 に限定したことから、受講希望者が多い講義は抽選によって履修を可能とせざるを得なかった。「統計入門」では特に文系向けのクラスで受講可能な数の 2 倍の学生が申し込み、大量の学生が受講できない事態となった。このような場合以外にも、時間割の編成上どうしても受講できない学生も出てくる可能性があり、そのような学生がデータ科学に触れる機会を作るために「統計の入門」は有効活用できると考えられる。

参考文献

- [1] 京都大学国際高等教育院 自己点検・評価報告書 第 1 号 (平成 30 年)
- [2] 教養・共通教育の改善に向けて企画評価委員会 WG 報告集、京都大学国際高等教育院 (平成 26 年)

3. シラバスの詳細

3.1 「統計入門」開講(平成 27 年度)

データサイエンス教育の充実の一步として「統計入門」は平成 27 年度に全学共通科目として設計・導入された。平成 29 年度までの 3 年間は、情報学研究科の教員 6 名が担当し 6 クラス開講し、各年度の履修者数は 800 名程度であった。当初は振れ幅の小さいシラバスに基づいた、統一された講義形式の実施が可能で、統制の取れた改善活動が可能であった。一方で、大規模クラスの存在と 1 クラス内での数学を中心とした受講生の基礎学力のばらつきという課題が生じた。

授業の概要・目的

統計に関する知識は、実験、試験、調査などの結果を用いた実証研究を行う上でなくてはならないものである。生活に関わるさまざまな効果やリスクがデータとともに語られ、生活者としても統計に対するリテラシーが求められるようになった。企業活動では、情報技術の発展によって、日々膨大なデータが生成されており、その活用が求められるようになった。本講は、研究や、生活、社会・経済活動に不可欠な統計を、集計・分析し、理解する力を養うことを目的とする。

ただし、統計や統計学については、膨大な研究の蓄積が有り、その利用はきわめて多分野に亘る。しかも、各分野で独自の発展をとげている部分もあり、本講のみでそのすべてを扱うことは出来ない。したがって、本講では、統計ならびに統計学に関する基本的な考え方を中心に講義することで、より発展的な統計・統計学の学習への礎となることを目指す。

具体的には、二元分割表（ 2×2 クロス集計表）の独立性の検定と関連性の強さの推定を主な題材として、統計データの収集、チェック、集計、分析、結果の解釈という一連の過程について解説し、統計データの発生、仮説検定と推定の考え方に関する理解を深める。また、発展的な学習の基礎として、平均の差の検定、分散分析、相関と回帰分析について、その基礎を解説する。

なお本講は、統計分析手順の機械的な利用や解釈だけを講義するのではなく、その基礎となる考え方を学ぶことを目指している。しかし、統計学的命題について、厳密な数学的証明は避け、あくまで統計・統計学のエンドユーザーとして必要とされる直感的な理解を目指す。

到達目標

1. 調査や実験・試験によるデータ収集の作法を理解する
2. データの種類や性質に応じたデータ確認と要約ができる

3. 二元分割表の独立性の検定と関連の強さの推定を行い、結果を解釈できる。
4. 仮説検定や推定の原理を理解する
5. 統計や統計学的知識を正しく使うための留意点と倫理を知る
6. 統計・統計学の応用について幅広く知り、今後の学習につなげる

授業計画と内容

- 第1講 概要
- 第2講 量的データの確認と要約
- 第3講 質的データの確認と要約
- 第4講 二元分割表とカイ二乗検定
- 第5講 二元分割表とフィッシャーの正確検定
- 第6講 二元分割表のリスク比とオッズ比
- 第7講 二元分割表におけるリスク差の検定・推定
- 第8講 小テストおよびフィードバック
- 第9講 検定・推定と標本規模の関係
- 第10講 確率分布と極限定理
- 第11講 平均の差の検定
- 第12講 分散分析
- 第13講 相関
- 第14講 回帰

履修要件

主に文系の学生が高校で履修したレベルの数学の知識を必要とする。

3.2 センター教員担当開始(平成 30 年度)

平成 29 年度の文部科学省の「数理・データサイエンス教育拠点校」に指定されたことを受けて、国際高等教育院に附属データ科学イノベーション教育研究センターが設置され、開講クラスは **2 倍の 12 クラスに増加**させることに成功した。全学向けの広報などの成果もあり、**令和 2 年度実績では約 1,200 名弱まで履修者数を増加**させることに成功した。この拡大の過程では、文系入試を経て入学した受講生も増えるなど、**数学を中心とした受講生の基礎学力のばらつきが拡大**するとともに、**担当教員の専門的背景も医学・農学・教育学・理学・工学と拡大**することとなった。

規模の拡大と専門性の拡大とともに、自由度を確保する必要性も生じたため、前述の通り、「授業計画と内容」を 14 区分から 9 区分に変更したうえで、**順番も明示しないよう変更**することになった。

授業の概要・目的

統計に関する知識は、実験、試験、調査などの結果を用いた実証研究を行う上でなくてはならないものである。生活に関わるさまざまな効果やリスクがデータとともに語られ、生活者としても統計に対するリテラシーが求められるようになった。企業活動では、情報技術の発展によって、日々膨大なデータが生成されており、その活用が求められるようになった。本講は、研究や、生活、社会・経済活動に不可欠な統計を、集計・分析し、理解する力を養うことを目的とする。

ただし、統計や統計学については、膨大な研究の蓄積が有り、その利用はきわめて多分野に亘る。しかも、各分野で独自の発展をとげている部分もあり、本講のみでそのすべてを扱うことは出来ない。したがって、本講では、統計ならびに統計学に関する基本的な考え方を中心に講義することで、より発展的な統計・統計学の学習への礎となることを目指す。

具体的には、二元分割表（ 2×2 クロス集計表）の独立性の検定と関連性の強さの推定を主な題材として、統計データの収集、チェック、集計、分析、結果の解釈という一連の過程について解説し、統計データの発生、仮説検定と推定の考え方に関する理解を深める。

なお本講は、統計分析手順の機械的な利用や解釈だけを講義するのではなく、その基礎となる考え方を学ぶことを目指している。しかし、統計学的命題について、厳密な数学的証明は避け、あくまで統計・統計学のエンドユーザーとして必要とされる直感的な理解を目指す。

到達目標

1. 調査や実験・試験によるデータ収集の作法を理解する
2. データの種類や性質に応じたデータ確認と要約ができる
3. 二元分割表の独立性の検定と関連の強さの推定を行い、結果を解釈できる。
4. 仮説検定や推定の原理を理解する
5. 統計や統計学的知識を正しく使うための留意点と倫理を知る
6. 統計・統計学の応用について幅広く知り、今後の学習につなげる

授業計画と内容

1. 概要と導入
2. 量的データの確認と要約
3. 質的データの確認と要約
4. JMP 等によるデータの確認と要約
5. 二元分割表とカイ二乗検定
6. 二元分割表とフィッシャーの正確検定
7. さまざまな確率分布と統計的検定の考え方
8. 中間のまとめと補足、小テストとフィードバック
9. 二元分割表のリスク比とオッズ比
10. 二元分割表におけるリスク差
11. 中心極限定理、区間推定の考え方
12. t 分布、検定・推定と標本規模
13. 統計と統計学の利用
14. 統・統計入門への繋がり

なお、講義の進度を反映して内容順序の変更や省略・追加を行うことがある。

履修要件

主に文系の学生が高校で履修したレベルの数学の知識を必要とする。

3.3 令和3年度のシラバス

「統計入門」を数理・データサイエンス・AI 教育プログラム認定制度のリテラシーレベルのプログラム科目として申請することとなり、令和3年度からは「**文科省のモデルカリキュラム等を反映して内容順序の変更や省略・追加を行うことがある**」旨の追記を行うこととなった。

授業の概要・目的 **変更なし**

統計に関する知識は、実験、試験、調査などの結果を用いた実証研究を行う上でなくてはならないものである。生活に関わるさまざまな効果やリスクがデータとともに語られ、生活者としても統計に対するリテラシーが求められるようになった。企業活動では、情報技術の発展によって、日々膨大なデータが生成されており、その活用が求められるようになった。本講は、研究や、生活、社会・経済活動に不可欠な統計を、集計・分析し、理解する力を養うことを目的とする。

ただし、統計や統計学については、膨大な研究の蓄積が有り、その利用はきわめて多分野に亘る。しかも、各分野で独自の発展をとげている部分もあり、本講のみでそのすべてを扱うことは出来ない。したがって、本講では、統計ならびに統計学に関する基本的な考え方を中心に講義することで、より発展的な統計・統計学の学習への礎となることを目指す。

具体的には、二元分割表（ 2×2 クロス集計表）の独立性の検定と関連性の強さの推定を主な題材として、統計データの収集、チェック、集計、分析、結果の解釈という一連の過程について解説し、統計データの発生、仮説検定と推定の考え方に関する理解を深める。

なお本講は、統計分析手順の機械的な利用や解釈だけを講義するのではなく、その基礎となる考え方を学ぶことを目指している。しかし、統計学的命題について、厳密な数学的証明は避け、あくまで統計・統計学のエンドユーザーとして必要とされる直感的な理解を目指す。

到達目標 **変更なし**

1. 調査や実験・試験によるデータ収集の作法を理解する
2. データの種類や性質に応じたデータ確認と要約ができる
3. 二元分割表の独立性の検定と関連の強さの推定を行い、結果を解釈できる
4. 仮説検定や推定の原理を理解する
5. 統計や統計学的知識を正しく使うための留意点と倫理を知る
6. 統計・統計学の応用について幅広く知り、今後の学習につなげる

授業計画と内容

14 区分から 9 区分に変更し、順番を明示しないよう変更

文科省のモデルカリキュラムへの言及を追加

- 概要と導入（1 回）
- データの確認と要約（2～3 回）
- 二元分割表と検定（2～3 回）
- さまざまな確率分布と統計的検定の考え方（1～2 回）
- 二元分割表のリスク比・オッズ比・リスク差（1～2 回）
- 中心極限定理、区間推定（1～2 回）
- t 分布、検定・推定と標本規模（1～2 回）
- 統計と統計学の利用（1 回）
- 発展的内容（1 回）

授業回数はフィードバックを含め全 15 回とする。

なお、講義の進度・文科省のモデルカリキュラム等を反映して内容順序の変更や省略・追加を行うことがある。

履修要件 変更なし

主に文系の学生が高校で履修したレベルの数学の知識を必要とする。

3.4 各担当教員の担当クラスの受講者と専門性を反映させる工夫

医学部医学科向けに開講されているクラスでは医学部学生の興味や専門性を鑑みて次のような工夫を行うことで学修効果改善を図っている。

- 最新の医学の学術雑誌論文の例などを示し、統計学の知識が実際に最先端の医学研究で活用されており、多くの受講者にとって将来的にも必要な学問であることを明示。
- 2元分割表、仮説検定や区間推定などを説明する際の様々な例において医学に関連するデータを採用することで、医学部の学生により強い関心を持ってもらう様に工夫。
- 医学部医学科の学生が卒業時に必ず受験する医師国家試験の出題範囲を示すことで学習意欲を高める様に工夫。
- 「医学教育コアカリキュラム」の中の「B 社会と医学・医療」の多くが、「統計入門」で扱われていることを説明し、受講意欲を高める様に工夫。
- (特に入学間もない)医学部医学科の学生は数学にも強い関心を持っているため、高度な質問に関して回答動画、資料などで対応。

理学部向けに開講されているクラスでは、主に数学・物理のような理論系の分野を指向する学生と主に化学・生物のような実験系の分野を志向する学生が並存することを鑑みて以下のような工夫を行うことで学習効果改善を図っている。

- 数学科目の確率論基礎と内容が重複する部分があるが、データを使うことを重要視した説明に特化し、差別化を図っている。
- 定理の証明はしない一方で理論的な理解の機会を損なわないよう注意し、具体例や図による可視化を用いることで理論を直感的に理解する様に工夫している。
- 現代におけるデータを扱うシーンでは統計ソフトウェアの利用が不可欠であり、講義全編を通じて R で学ぶことができるようにコンテンツを配布している。

文系学部向けに開講されているクラスでは、数理系に必ずしも強くない学生が多いことに配慮して、次のような工夫を行うことで学習効果改善を図っている。

- 統計的手法の数理的な考察は最小限にとどめ、各手法の使い所、直感的な意味を平易な言葉で時間をかけて解説する一方で、学ぶべきコンテンツの量が他のクラスと比べて少なくなならないように工夫している。
- 仮説検定や信頼区間のような、大学で新しく学習する統計的な考え方のロジックは、様々なシチュエーションで繰り返し学べるように配慮をしている。
- 近年では、文系分野でもデータサイエンスが多くの知識発見に貢献しているこ

とを強調し、学習意欲を高めるように工夫している。

- 社会科学、教育学的なデータ例を多く用いることで、次の学びへの動機付けとなるように工夫している。

工学部向けに開講されているクラスでは、工学部学生の興味や専門性を考慮し、次のような工夫を行うことで学習効果改善を図っている。

- 実社会での統計の活用を意識した内容を積極的に取り入れ、工学と同じように実社会で役立つ点を強調し、工学部学生の学習意欲を高める工夫をしている。
- 2 元分割表やリスク比、オッズ比など工学では必ずしも主流でない内容については、実生活との関わりが深い事例を多数挙げ、学生の興味を引く工夫をしている。
- 上記に関連して、学生に向けて簡単なアンケート調査を初回授業時に実施し、そのデータの要約・可視化・仮説検定に関する演習を実施している。
- 正規分布に関連した事柄は、工学部学生の数学レベルを考慮しながら、数式を使ったより詳しい解説を行い、より深い理解が得られるよう工夫している。また視覚的で直感的な説明を用意し、学生の理解を促進する工夫をしている。

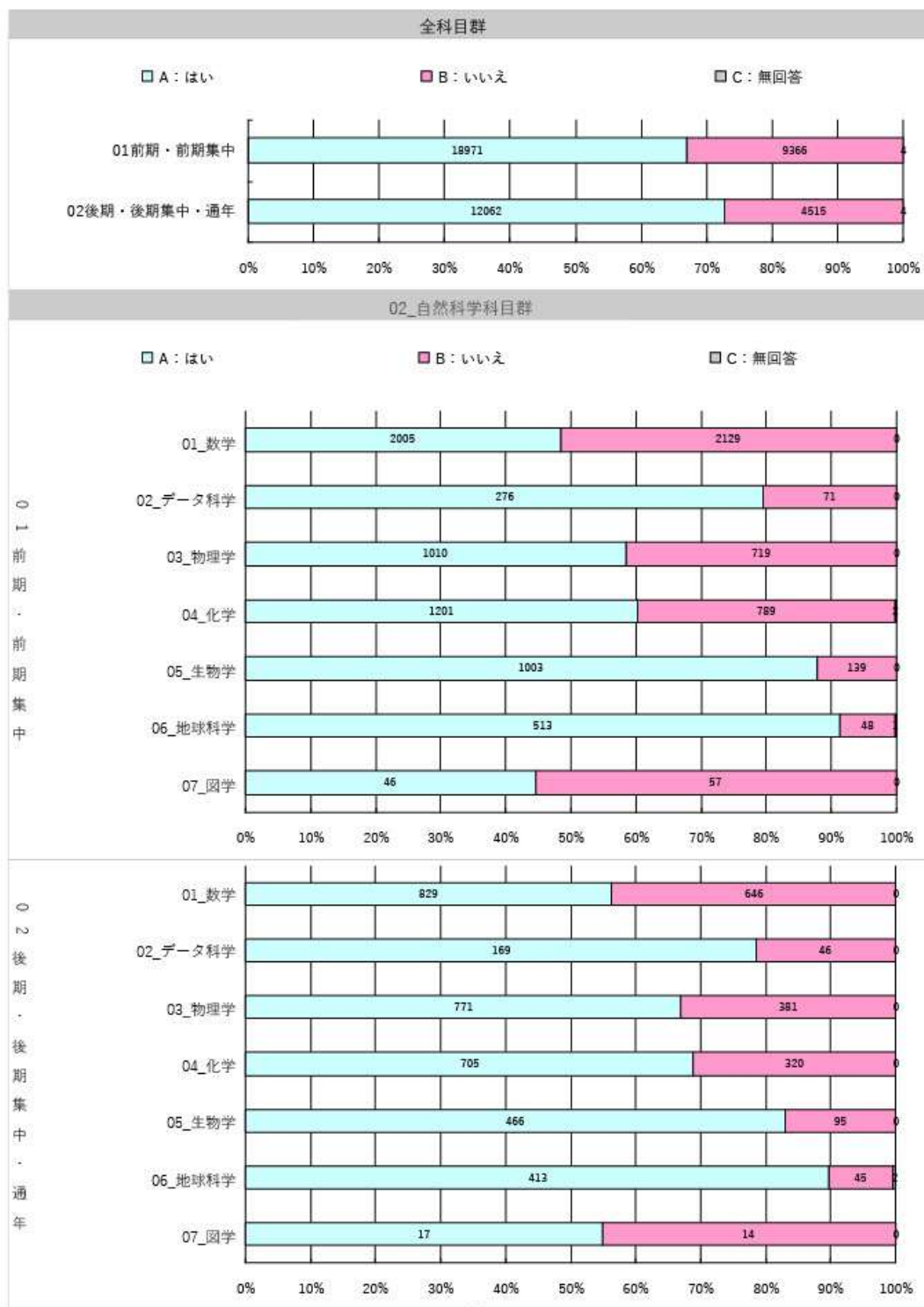
3.5 シラバスの有効性

全学共通科目では、すべての科目について講義の最終週に学生へのアンケートを実施している。その結果を見ると、「統計入門」を含むデータ科学群科目については、「シラバスは有効であった」と回答する学生が、数学や物理に比して多いことが判明した(次ページ図参照)。

原因として考えられるのは、数学や物理などは理系学部・学科では必須あるいは履修が推奨されており、学生にとっても高校時代から馴染みがあり、大学での講義の内容や方向性も想像しやすい一方で、データ科学群科目には馴染みが少なく、シラバスで内容を知ってから履修や学習の計画を立てているのではないか、ということである。この点は検証が必要であるが、令和3年度に e-learning「統計の入門」を5月の連休中に済ませておくように指示したクラスの学生からは、「今後の講義の展開がわかって有用であった」という反応があったことから伺い知ることができる。

令和2年度授業アンケート結果より

前期Q3（後期Q6） 履修登録の際にこの授業のシラバスを活用しましたか。



4. 関連科目

いわゆる文系入試を経てきた学生が「統計入門」をはじめとするデータサイエンス科目を履修するときに必要な知識を補完するために平成 30 年度から開講している科目である。1 回生に対しては数学の講義を学ぶ時の動機づけになるように、2 回生以上に対しては学部専門科目の中でデータサイエンスを学ぶ際に 1 回生の数学の内容をデータサイエンスという視点から復習できるように構成している。さらに、高等学校での数学における統計的内容を復習することにより、高等学校の教育から大学での専門基礎教育へ円滑に移行するための橋渡しとしての機能も持たせている。

4.1 数理データサイエンスのための数学入門 I

授業の概要・目的

高度情報化社会である今日、至るところに蓄積される大量のデータを解析するための科学であるデータ科学は、学術全般・産業界のみならず日常生活の至る所に大きな変化をもたらそうとしている。データ科学の根幹である情報学・統計学・数理科学に対する基本的な理解、特に基礎的な数学の素養は社会を支える広範な人材にとっての基礎的な教養となりつつある。

本講義は、データ科学における理論と技術に必要な基礎数学の中で、基礎的な確率・統計および基礎的な線形代数について、データ解析への応用と並行しながら理解することを目的とする。数学理論としての完全な体系よりも、具体的な計算手法の習得とデータ解析への応用方法の理解を重視する。また、文系学生が受講可能なように、高校での数学 III の知識を仮定せず、必要が生じれば、その都度補う形で進める。

到達目標

確率・統計および線形代数がどのようにデータ解析に用いられるかを理解する。数学的理論だけでなく、具体的な計算方法について習得する。

授業計画と内容

次の内容について 15 週で講述する予定である。

1. 導入：確率・統計および線形代数の応用による、大量のデータの背後に隠れた規則性・知識の抽出（1 週）
2. 条件付確率の応用（3 週）
条件付き確率の情報学的意味、条件付き確率の繰返し適用、MAP 推定と応用
3. 統計の基礎（4 週）
（平均と分散、確率変数と確率分布、二項分布、正規分布とその利用）

4. 多変量のデータ分析の基礎 (2 週)

(同時確率分布、周辺分布、共分散と相関係数)

5. データ解析の線形代数の基礎 (5 週)

(回帰、行列と連立一次方程式の解法、主成分、Lagrange の未定乗数法、行列の固有値、因子分析、行列の積)

講義の進度、受講者の理解度によって回数の割り当てと内容の変更がありうる。

履修要件

高等学校での数学 A (場合の数と確率)、数学 I (データの整理)、数学 B (平面上と空間のベクトル) を学習していること。

数学 B (確率分布と統計的な推測) については講義中で扱う。

4.2 続・統計入門

3.2 で述べたとおり、平成 30 年度より統計入門のシラバスを一部変更した。それは、履修者や教員の特性に合わせて、内容に幅を持たせるためでもあるが、平成 29 年度以前の統計入門は内容的に詰め込み過ぎであるという指摘があったことにも依る。そのため、従来の統計入門では講述していたが、新しい統計入門では省略されうる内容に加えて、因果推論など少し発展的な話題を取り入れた「続・統計入門」を新設し、前期と後期に 1 クラスずつ開講した。

授業の概要・目的

統計に関する知識は、実験、試験、調査などの結果を用いた実証研究を行う上でなくてはならないものである。生活に関わるさまざまな効果やリスクがデータとともに語られ、生活者としても統計に対するリテラシーが求められるようになった。企業活動では、情報技術の発展によって、日々膨大なデータが生成されており、その活用が求められるようになった。本講は、研究や、生活、社会・経済活動に不可欠な統計を、集計・分析し、理解する力を養うことを目的とする。

ただし、統計や統計学については、膨大な研究の蓄積が有り、その利用はきわめて多分野に亘る。しかも、各分野で独自の発展をとげている部分もあり、本講のみでそのすべてを扱うことは出来ない。本講では科目「統計入門」で扱えなかった、やや発展的な話題を中心に講義することで、より発展的な統計・統計学の学習への礎となることを目指す。

具体的には、平均の差の検定、分散分析、相関と回帰分析の基礎について解説するとともに、一般化線形モデルによるこれらの統一的な理解を行う。さらに、因果推論の基本的な考え方と、具体的な方法についても解説を行う。

なお本講は、統計分析手順の機械的な利用や解釈だけを講義するのではなく、その基礎となる考え方を学ぶことを目指している。しかし、統計学的命題について、厳密な数学的証明は避け、あくまで統計・統計学のエンドユーザーとして必要とされる直感的な理解を目指す。

到達目標

1. 統計的検定と推定の考え方を理解し、これを実施できる。
2. 平均の差の検定・分散分析の考え方を理解し、これを実施できる。
3. 相関と回帰について理解し、これを実施できる。
4. 因果推論の基本的な考え方と手法を理解する。
5. 統計・統計学の手法と応用について幅広く知り、今後の学習につなげる。

授業計画と内容

第 1 講 概要と導入

第 2～4 講 検定と推定

第 5～6 講 分散分析

第 7 講 相関

第 8 講 中間テストとフィードバック

第 9～10 講 回帰分析

第 11～12 講 因果推論の基礎

第 13～14 講 発展的事項とまとめ

(上記予定は目安であり、実際の講義の進度に応じて変更・前後することがある)

履修要件

「統計入門」レベルの内容を理解していることが望ましい。

4.3 統計と人工知能

「続・統計入門」は履修登録者数がそれほど多くなく、「統計入門」では飽き足らない学生に対しその次の科目を提供する、という役割は果たしているが、リテラシーとして多くの学生が受講したとは言いにくい状態であった。ちょうど令和元年に「AI戦略2019」が内閣府・文部科学省から発表されたことを受け、令和2年からは学生自体が強く興味を惹かれるであろう人工知能の内容を増強し「統計と人工知能」という名前に改称し、引き続き前期と後期に1クラスずつ開講した。

授業の概要・目的

統計に関する知識は、実験、試験、調査などの結果を用いた実証研究を行う上でなくてはならないものである。生活に関わるさまざまな効果やリスクがデータとともに語られ、生活者としても統計に対するリテラシーが求められるようになった。企業活動では、情報技術の発展によって、日々膨大なデータが生成されており、その活用が求められるようになった。

本講は、研究や、生活、社会・経済活動に不可欠な統計を、集計・分析し、理解する力を養うことをひとつめの目的として、科目「統計入門」で扱えなかった、やや発展的な話題を中心に講義することで、より発展的な統計・統計学の学習への礎となることを目指す。具体的には、平均の差の検定、分散分析、相関と回帰・判別分析の基礎について解説するとともに、一般化線形モデルによるこれらの統一的な理解を行う。さらに、因果推論の基本的な考え方と、具体的な方法についても解説を行う。

さらに、本講では第二の目的として、近年注目されている人工知能について、その基本的な考え方や応用について理解することで、現在も目まぐるしい早さで発展している当該分野への橋渡しを行う。近年の人工知能技術は膨大なデータをもとにした統計的なアプローチに基づいており、統計学がどのような形でその基礎を形作っているかを中心に学ぶ。具体的には、機械学習における基本的な統計的手法をはじめ、ニューラルネットワーク、深層学習など、近年大きく発展している技術、また、それらの応用として自然言語処理や画像処理などの話題について解説する。

本講は、統計分析手順や人工知能手法の機械的な利用や解釈だけを講義するのではなく、その基礎となる考え方を学ぶことを目指しているが、あくまで直感的な理解を目指すことを主な目標とし、厳密な数学的証明等は避ける。

到達目標

1. 統計的検定と推定の考え方、とくに平均の差の検定・分散分析の考え方を理解し、これを実施できる

2. 相関と回帰について理解し、これを実施できる
3. 因果推論の基本的な考え方と手法を理解する
4. 人工知能・機械学習の基本的な考え方と手法を理解する
5. 統計と人工知能の手法と応用について幅広く知り、今後の学習につなげる

授業計画と内容

授業回数はフィードバックを含め全 15 回とする

- 統計の基礎概念の復習（検定・推定）3～4 回
 - 相関・回帰分析 2 回
 - 因果推論 2 回
 - 人工知能概説 1 回
 - 機械学習 2 回
 - ニューラルネットワーク・深層学習 1～2 回
 - 自然言語処理・画像処理 1～2 回
 - 発展的課題 1 回
- （上記予定は目安であり、実際の講義の進度に応じて変更・前後することがある）

履修要件

「統計入門」レベルの内容を理解していることが望ましい。

5. e-learning 教材 (KoALA・gacco・放送大学)

2.4 で述べたように、「統計入門」担当の教員が増加したことにより、学内的には2.1 で述べた「統計入門」の設計段階での実施方針(2)にある e-learning コンテンツの整備の必要性が指摘されはじめた。同時に、数理・データサイエンス教育拠点校として、協力校や連携校を中心に、数理・データサイエンス教育の充実を目指す大学に向けて、教育コンテンツの提供をすすめる必要性も生じるようになった。

これを受けて、センターとしては、高等教育研究開発推進センターがオンデマンドコンテンツを提供している Kyoto University Online for Augmented Learning Activities (KoALA)上に、「統計の入門」を展開し、毎学期開講し「統計入門」などの補助教材として活用している。この「統計の入門」は「統計入門」のポイントを圧縮したもので、著作権処理後のスライドをさらにプロの手を借りて閲覧性を向上させ、さらにプロのナレーターによる音声を付与したものである。



KoALA 講義を探す 登録 サインイン

統計の入門 KOALA

DS004 に登録登録

講義の概要

統計に関する知識は、実験、試験、調査などの結果を用いた実証研究を行う上でなくてはならないものである。生活に関わるさまざまな効果やリスクがデータとともに語られ、生活者としても統計に対するリテラシーが求められるようになった。企業活動では、情報技術の発展によって、日々膨大なデータが生成されており、その活用が求められるようになった。本講は、研究や、生活、社会・経済活動に不可欠な統計を、集計・分析し、理解する力を養うことを目的とした「統計入門」「統計入門」を圧縮した内容になっている。これから統計を学ぶとする初学者や、学び直しを目指す学生を主たる対象としている。

本講では、統計ならびに統計学に関する基本的な考え方を中心に講義すること。より発展的な統計・統計学の学習への礎となることを目指す。統計分析手順の機械的な利用や解釈というよりも、その基礎となる考え方を学ぶことを目指している。また、厳密な数学的証明は避け、あくまで統計・統計学のエンドユーザーとして必要とされる直感的な理解を目指している。

- ・データ科学における統計学について
- ・統計学を学ぶ意義について
- ・統計処理に用いられるソフトウェアについて
- ・データの特性について
- ・データの可視化について
- ・2元分割表について
- ・検定と推定の考え方について
- ・相関と回帰の考え方について
- ・因果推論の考え方について

学習の内容を見る

講師

京都大学国際高等教育院データ科学教室

https://koala.highedu.kyoto-u.ac.jp/courses/coursev1:KoALA+DS004+2021_T2/about
より転載

「統計の入門」を KoALA 上で初めて開講した令和 2 年度の新学期は、折しも COVID-19 感染拡大にすべての大学が苦慮している時期となり、高等教育研究開発推進センターには学内からのアクセスだけでなく協力校や連携校からのアクセスも可能な設定にさせていただくべく配慮をいただいた。その結果、学内だけでなく他大学からも含めたアクセスは 750 名となり、実績が蓄積される中で不具合箇所の改修もすすめることができた。そうした改善のプロセスの産物として、令和 2 年度には一般社団法人日本オープンオンライン教育推進協議会(JMOOC)を通じてドコモ社が提供する MOOC システムである「gacco」上にも展開し、京都大学初のコンテンツを JMOOC に提供することも実現できた。全国からの受講者は 11,539 名にのぼり、その中には社会人も多く含まれる。


[登録](#)
[ログイン](#)

統計の入門

受講登録は終了しました



講座概要

講座内容

統計に関する知識は、実験、試験、調査などの結果を用いた実証研究を行う上でなくてはならないものである。生活に関わるさまざまな効果やリスクがデータとともに語られ、生活者としても統計に対するリテラシーが求められるようになった。企業活動では、情報技術の発展によって、日々膨大なデータが生成されており、その活用が求められるようになった。本講座は、研究や、生活、社会・経済活動に不可欠な統計を、集計・分析し、理解する力を養うことを目的とした「統計入門」「統計入門」を圧縮した内容になっている。これから統計を学ぼうとする初學者や、学び直しを目指す学生を主たる対象としている。

本講座では、統計ならびに統計学に関する基本的な考え方を中心に講義することで、より発展的な統計・統計学の学習への礎となることを目指す。統計分析手順の機械的な利用や解釈というよりも、その基礎となる考え方を学ぶことを目指している。また、厳密な数学的証明は避け、あくまで統計・統計学のエンドユーザーとして必要とされる内容の直感的な理解を目指している。

第1章：イントロダクション

第2章：データ特性・可視化

第3章：統計ソフト

第4章：2元分割表

第5章：検定・推定

第6章：相関と回帰

第7章：因果推論とまとめ

講師・スタッフ紹介

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



※受講登録するとお客様の利用情報は講座提供者（講師）に共有されます。詳しくは[利用規約](#)と[プライバシーポリシー](#)をご覧ください。

講座番号: ga150

受講開始日: 2021年1月26日

想定される勉強時間/週: 3時間程度

https://lms.gacco.org/courses/course-v1:gacco+ga150+2021_01/about

より転載

このようにセンターとしては「統計入門」の充実と発展に注力してきたが、大学入学直後の学生向けの導入となるリテラシーレベルの統計関連講座の充実、拠点校・協力校・連携校の中でも特徴的な訴求点の一つとなっている。拠点校を中心として形成されている「数理・データサイエンス教育強化拠点コンソーシアム」では、2年に及んだ議論を経て、「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～」を策定し公表している。（下図参照）

http://www.mi.u-tokyo.ac.jp/consortium/pdf/model_literacy.pdf

● モデルカリキュラムと教育方法

導入	1. 社会におけるデータ・AI活用		● データ・AI活用事例を紹介した動画（MOOC等）を使った 反転学習 を取り入れ、講義ではデータ・AI活用領域の広がりや、技術概要の解説を行うことが望ましい。 ● 学生がデータ・AI活用事例を調査し発表する グループワーク 等を行い、一方通行で事例を話すだけの講義にしないことが望ましい。
	1-1. 社会で起きている変化	1-2. 社会で活用されているデータ	
基礎	1-3. データ・AIの活用領域	1-4. データ・AI活用のための技術	● 各大学・高専の特徴に応じて 適切なテーマ を設定し、 実データ （あるいは模擬データ）を用いた講義を行うことが望ましい。 ● 実際に手を動かしてデータを可視化する等、学生自身がデータ活用プロセスの一部を 体験 できることが望ましい。 ● 必要に応じて、フォローアップ講義（ 補講 等）を準備することが望ましい。
	1-5. データ・AI活用の現場	1-6. データ・AI活用の最新動向	
心得	2. データリテラシー		● データ駆動型社会のリスクを 自分ごと として考えさせることが望ましい。 ● データ・AIが引き起こす課題について グループディスカッション 等を行い、一方通行で事例を話すだけの講義にしないことが望ましい。
	2-1. データを読む	2-2. データを説明する	
選択	3. データ・AI活用における留意事項		● 本内容は オプション 扱いとし、大学・高専の特徴に応じて学修内容を選択する。 ● 各大学・高専の特徴に応じて 適切なテーマ を設定し、 実データ （あるいは模擬データ）を用いた講義を行うことが望ましい。 ● 学生が希望すれば本内容を受講できるようにしておくことが望ましい（ 大学間連携 等）。
	3-1. データ・AIを扱う上での留意事項	3-2. データを守る上での留意事項	
選択	4. オプション		
	4-1. 統計および数理基礎	4-2. アルゴリズム基礎	
選択	4-3. データ構造とプログラミング基礎	4-4. 時系列データ解析	
	4-5. テキスト解析	4-6. 画像解析	
選択	4-7. データハンドリング	4-8. データ活用実践（教師あり学習）	
	4-9. データ活用実践（教師なし学習）		

全国展開のためには、これに基づいた教材も必須であるため、実績豊富な放送大学と連携して制作することとなったが、特に体験も含めた重要なパートである「基礎」に関しては、十分な経験を有する大学が乏しく、全8回のすべてを京都大学の教員が分担して担当することとなった。（下図参照）

数理・データサイエンス・AI リテラシー講座 基礎

放送回	副題	担当講師 ※所属職名は4月1日時点	放送日時	
第1回	データを読む1 結論を導く作法	京都大学 山本教授	4/29（木）6:00	4/30（金）21:45
第2回	データを読む2 量的データの要約	京都大学 林教授	4/30（金）6:00	5/1（土）21:45
第3回	データを読む3 2変数のデータの相関関係	京都大学 中野特定講師	5/6（木）6:00	5/7（金）21:45
第4回	データを説明する1 データを正しく読む/質的データの分析	京都大学 鹿島教授	5/7（金）6:00	5/8（土）21:45
第5回	データを説明する2 データの図表表現	京都大学 木村特定講師	5/13（木）6:00	5/14（金）21:45
第6回	データを説明する3 データの比較と可視化	京都大学 関戸特定講師	5/14（金）6:00	5/14（土）21:45
第7回	データを扱う1 EXCELでデータサイエンス!: 基本操作編	京都大学 田村教授	5/20（木）6:00	5/21（金）21:45
第8回	データを扱う2 EXCELでデータサイエンス!: 実践編	京都大学 田村教授	5/21（金）6:00	5/22（土）21:45

これらのコンテンツの大半は、「統計入門」で培われた経験がベースになっているが、高校数学などの学習が十分ではない対象者を意識したものであることもあり、放送大学の教員の助言も受けて、初学者にとって更に入りやすい教材として完成させることができた。2021 年度から 2024 年度まで、放送大学の学生のみならず BS231 チャンネルにより、全国民向けに繰り返し放映される予定である。

6. 受講促進活動

令和 2 年度より、学部個別学力試験の合格発表直後に学部・大学院の新入学生全員に送付される宅配便にデータサイエンス科目の履修を促すフライヤーを封入している。

さらに、学部新入生全員に対する新入生ガイダンスにおいて、「ビッグデータの時代 データ科学を学ぼう」という題目でデータサイエンス科目の履修を促す説明を行うとともに（令和 3 年度）、さらに詳細な動画を公開している（令和 2 年度・3 年度）。

資料編

「統計入門」履修者実績

学部・学科名称	収容定員	令和2年度		令和元年度		平成30年度		平成29年度		平成28年度		平成27年度		履修者数合計	履修率
		履修者数	合格者数	履修者数	合格者数	履修者数	合格者数	履修者数	合格者数	履修者数	合格者数	履修者数	合格者数		
総合人間学部	480	90	54	64	43	61	37	52	25	-	-	-	-	267	56%
文学部	880	76	44	41	24	34	23	18	7	-	-	-	-	169	19%
教育学部	260	48	39	45	36	49	37	30	23	-	-	-	-	172	66%
法学部	1,340	42	28	49	26	33	22	15	6	-	-	-	-	139	10%
経済学部	1,000	34	19	31	10	17	6	11	5	-	-	-	-	93	9%
理学部	1,244	263	182	192	104	205	115	179	85	-	-	-	-	839	67%
医学部 医学科	642	112	108	111	107	115	112	119	111	123	115	254	222	834	130%
医学部 人間健康科学科	451	50	45	29	20	28	22	46	29	-	-	-	-	153	34%
薬学部	380	85	46	38	26	86	68	-	-	-	-	-	-	209	55%
薬学部 薬学科	135	-	-	-	-	-	-	27	24	24	15	29	29	80	59%
薬学部 薬科学科	245	-	-	-	-	-	-	49	37	-	-	-	-	49	20%
工学部	3,820	333	222	367	241	297	185	240	142	-	-	-	-	1,237	32%
農学部	1,200	44	28	37	20	23	10	11	6	-	-	-	-	115	10%
合 計	11,697	1,177	815	1,004	657	948	637	797	500	147	130	283	251	4,356	37%

注1) 医学部医学科及び薬学部薬学科は6年制

注2) 医学部医学科の履修率が100%を超えているが、これは本プログラムの科目「統計入門」が導入された平成27年度に配当学年以外の学生の履修が多かったためである

注3) 薬学部はH30より4年次に学科振分けを行うこととなった。

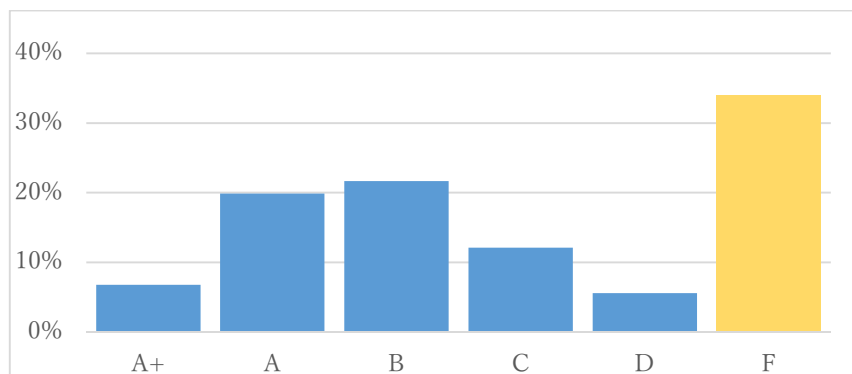
注4) 収容定員は2020年度現在の収容定員数。

注5) 履修者は履修確定当初人数。履修取消や中途退学者等は反映していない。

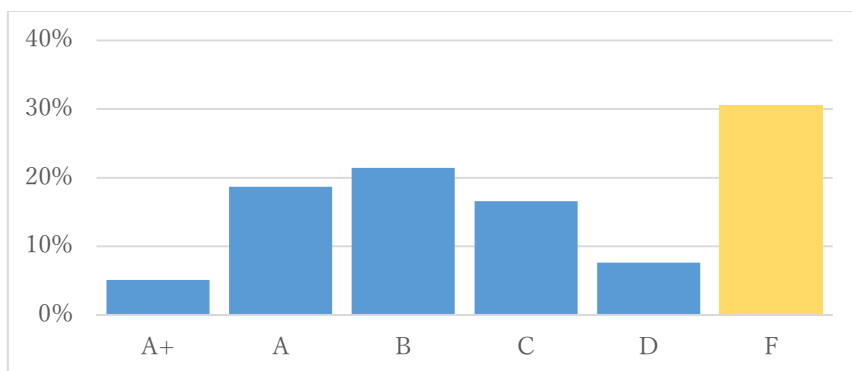
年度別成績分布

教育効果検証のため履修者数は履修取消や中途退学者等を除外し、履修者数に対する成績別の割合を示す

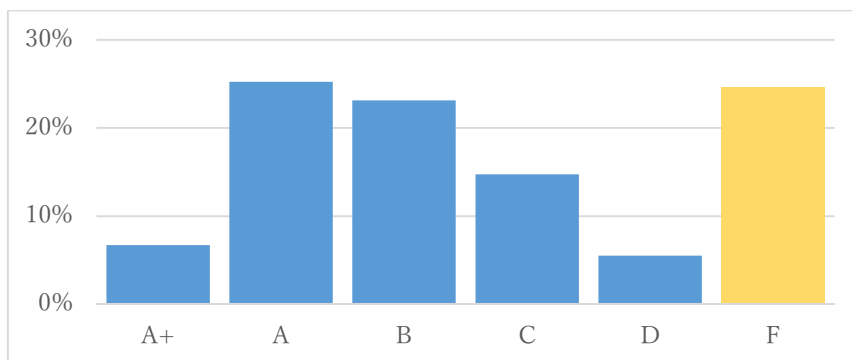
平成 28 年度



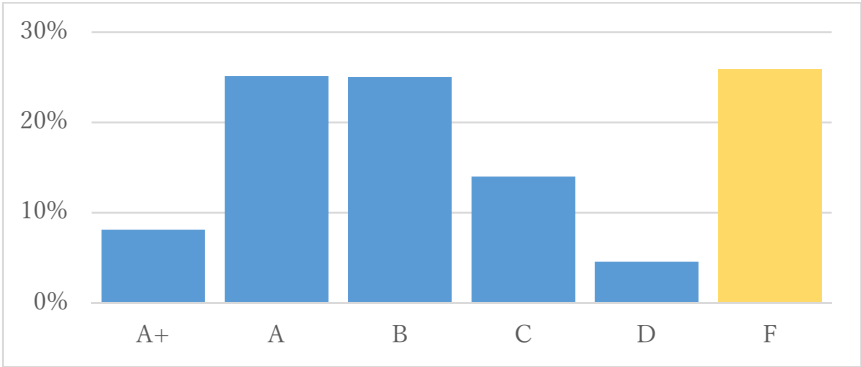
平成 29 年度



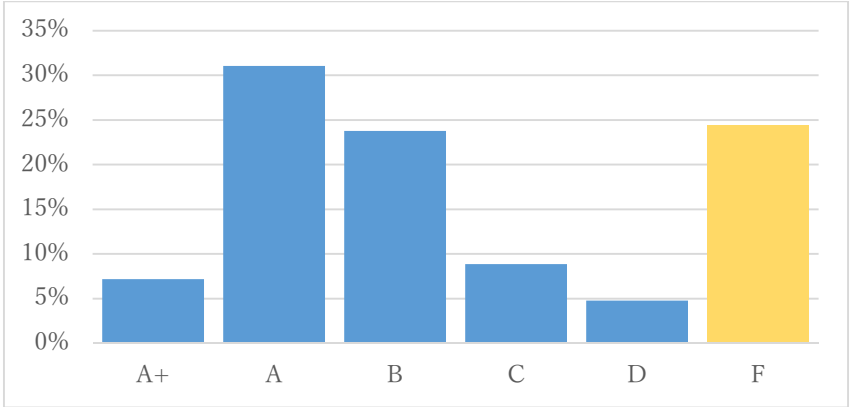
平成 30 年度



平成 31(令和元)年度



平成 31(令和 2)年度



クラス別成績の詳細

開 講 期	文系向等	履修者数 (取消除)	A+	A	B	C	D	合格	合格 率	A+,A の割合	A+,A,B の割合
平成 28 年度											
前		150	22	57	34	12	2	127	84.7%	62.2%	89.0%
前		76	14	22	14	3	1	54	71.1%	66.7%	92.6%
前		151	9	21	27	25	7	89	58.9%	33.7%	64.0%
後		169	2	25	58	15	1	101	59.8%	26.7%	84.2%
後		103	0	9	14	19	22	64	62.1%	14.1%	35.9%
後		85	3	12	12	15	8	50	58.8%	30.0%	54.0%
	合計	734	50	146	159	89	41	485	66.1%	26.7%	48.4%

平成 29 年度											
前		133	16	40	38	15	5	114	85.7%	49.1%	82.5%
前		70	4	16	13	11	3	47	67.1%	42.6%	70.2%
前		147	5	25	24	28	9	91	61.9%	33.0%	59.3%
後		127	5	12	22	20	9	68	53.5%	25.0%	57.4%
後		125	4	32	30	18	11	95	76.0%	37.9%	69.5%
後		121	3	10	28	28	18	87	71.9%	14.9%	47.1%
	合計	723	37	135	155	120	55	502	69.4%	23.8%	45.2%

平成 30 年度											
前		60	4	26	14	9	2	55	91.7%	54.5%	80.0%
前		59	7	25	14	8	3	57	96.6%	56.1%	80.7%
前		32	1	4	9	5	3	22	68.8%	22.7%	63.6%
前		79	0	17	16	10	11	54	68.4%	31.5%	61.1%
前		62	6	22	19	4	0	51	82.3%	54.9%	92.2%
前		172	8	46	52	26	9	141	82.0%	38.3%	75.2%
前	指定無	82	12	23	14	8	2	59	72.0%	59.3%	83.1%
前	文系向け	58	3	7	6	15	6	37	63.8%	27.0%	43.2%
後		92	3	16	34	17	5	75	81.5%	25.3%	70.7%
後		88	10	13	7	11	5	46	52.3%	50.0%	65.2%
後	文系向け	64	3	15	11	12	1	42	65.6%	42.9%	69.0%
	合計	848	57	214	196	125	47	639	75.4%	32.0%	55.1%

平成 31(令和元)年度											
前		65	14	16	16	12	1	59	90.8%	46.2%	70.8%
前		53	7	13	20	6	4	50	94.3%	37.7%	75.5%
前		32	2	7	7	3	3	22	68.8%	28.1%	50.0%
前		71	0	16	19	9	2	46	64.8%	22.5%	49.3%
前		68	8	19	10	7	1	45	66.2%	39.7%	54.4%
前		111	1	26	41	20	2	90	81.1%	24.3%	61.3%
前		107	2	46	24	11	9	92	86.0%	44.9%	67.3%
前	指定無	56	21	13	19	9	4	41	73.2%	60.7%	94.6%
前	文系向け	78	5	19	16	12	0	52	66.7%	30.8%	51.3%
後		58	1	17	16	7	3	44	75.9%	31.0%	58.6%
後		99	9	17	16	17	9	68	68.7%	26.3%	42.4%
後	文系向け	101	3	17	21	13	3	57	56.4%	19.8%	40.6%
	合計	899	73	226	225	126	41	666	74.1%	33.3%	58.3%

令和 2 年度											
前		56	3	20	19	6	3	51	91.1%	41.1%	75.0%
前		70	9	20	19	7	8	63	90.0%	41.4%	68.6%
前		56	0	14	19	10	3	46	82.1%	25.0%	58.9%
前		99	10	28	19	13	3	73	73.7%	38.4%	57.6%
前		93	4	43	20	5	3	75	80.6%	50.5%	72.0%
前		92	9	25	31	9	2	76	82.6%	37.0%	70.7%
前		95	8	49	17	2	3	79	83.2%	60.0%	77.9%
前	指定無	82	10	17	19	9	4	59	72.0%	32.9%	56.1%
前	文系向け	147	17	55	22	8	6	108	73.5%	49.0%	63.9%
後		106	1	24	31	6	3	65	61.3%	23.6%	52.8%
後		66	4	13	15	11	6	49	74.2%	25.8%	48.5%
後	文系向け	123	3	29	27	10	8	77	62.6%	26.0%	48.0%
	合計	1085	78	337	258	96	52	821	75.7%	38.2%	62.0%

注) 教育効果検証のため履修者数は履修取消や中途退学者等を除外している。

統計入門 第1回 概要

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
(情報学研科併任)
林 和則
hayashi.kazunori.4w@kyoto-u.ac.jp

例1: 政府統計

・我が国の人口・世帯の実態は？

- ・ 国勢調査(こくせいちょうさ):
 - ・ ある時点における人口及び、その性別や年齢、配偶の関係、就業の状態や世帯の構成といった人口及び世帯に関する各種属性のデータを調べる全数調査(Wikipedia)

・たとえば:

- ・ 産業別の就業者数はどのように年推移しているか?
- ・ 各地の昼間人口は?
- ・ 大規模で莫大なコストがかかる
 - ・ 費用コスト・時間コスト



統計の利用例

- ・ 政府統計
 - ・ 国勢調査・工業統計調査など
- ・ サンプル調査
 - ・ テレビの視聴率・内閣支持率・など
- ・ 薬効評価
 - ・ ある新薬に実際に効果があるかどうかを判断
- ・ Webサービス開発
 - ・ ある新機能に実際に効果があるかどうかを判断
- ・ マーケティング
 - ・ Webショッピングサイトでのおすすめ

様々な政府統計

・国勢調査をはじめ様々な

- ・ 国勢調査: 日本全国と地域別の人口・世帯とその内訳
- ・ 工業統計調査: 全国と地域別の工業の従業者数、出荷額など
- ・ 商業統計: 日本全国と地域別の商業の従業者数、販売額など
- ・ 農林業センサス: 日本全国と地域別の農家などの就業状態、農業生産、販売などの状況
- ・ 消費者物価指数: 商品やサービスの物価の変化
- ・ 労働力調査: 就業者や失業者数、その内訳(男女、年齢別など)
- ・ 家計調査: 日本中の家庭の家計の収入・支出の状況
- ・ 鉱工業生産指数: 日本中の工場などでの生産高の動き

公表統計

・政府統計の総合窓口 e-Stat

- ・ さまざまな政府統計の結果を公開
- ・ PSI(ポケット統計情報)なども利用可能



例2: サンプル調査

・少数の人のデータから国全体の傾向を知る

- ・ テレビの視聴率
 - ・ 100人に対する調査で6名が視聴していると回答
 - ・ 国民全員のうち実際に視聴しているのは何人か?
- ・ 内閣支持率
 - ・ 100人の道行く人に質問する
 - ・ 50人が内閣を支持すると回答
 - ・ 国民全員のうち支持するのは何人か?
- ・ あるいは、概ね何人から何人の間におさまっているか?
 - ・ どのくらいの信用性があるかが重要



例3: 薬は効くか

・新薬を投与したら回復率が上がった。この薬は効く?

- ・ 目的
 - ・ ある新薬に実際に効果があるかどうかを判断したい
- ・ 手順
 - ・ 薬を投与したグループとしなかったグループに分けて回復率の違いをみる
- ・ 結果
 - ・ 2つのグループの回復率を比較したところ、投与グループの回復率のほうが13%高かった
- ・ 判断
 - ・ この結果をもって「この新薬に効果がある」と言えるか?

例4: どちらのサービスがよい?

・A/Bテストでウェブサービスを最適化する

- ・ WebサービスやWeb広告のデザイン
 - ・ どちらのメニュー配置が使いやすいか?
 - ・ 赤と青、どちらが訴求力があるか?
 - ・ どちらの広告がクリックされやすいか?
- ・ 実際に試してみてもどちらがよいかを定める
 - ・ Web上のサービスなので効果測定が可能
- ・ 実際の例
 - ・ オプションAとオプションBのどちらかをユーザーに提示する。Aのクリック率は2%、Bのクリック率は1%であったときAはBよりも良いと言えるか?



例5: マーケティング

- 消費者の購買行動を予測する
 - Webショッピングサイトでのおすすめ商品
 - Amazon.co.jp の推薦機能
- 誰に何を薦めると買ってくれるだろうか?
 - これまでの購買履歴をもとに、ある商品を買ってくれるかどうか予測して、最も購買可能性が高いものから提示する
- データマイニング
 - 機械学習技術を用いて大量のデータから有益な情報を発見する



統計学の分類

- 統計には「記述統計」と「推測統計」がある
 - 記述統計**
 - 全数調査を前提
 - データを整理・視覚化することで理解
 - 推測統計**
 - サンプリング調査を前提
 - 部分から全体を知る
 - 仮説が正しいかを判断する
 - 過去から未来を予測する
 - 両者は確率論でつながる
 - 標本抽出時のゆらぎを考える



統計学の歴史

- 記述統計から推測統計へ
 - ～18世紀: 統計の誕生
 - 国家 (Staat) の実態をとらえるための統計
 - 今日の国勢調査などに対応
 - ラプラス: 古典確率論の成立
 - 19世紀: 記述統計の成立
 - 「近代統計学の父」ケトラー: 社会物理学
 - 「平均人」の概念・身体的データ (Body Mass Index: BMI)
 - ガウスによる誤差や正規分布の研究
 - 「ダーウィンの従弟」ゴルトン: 進化論 → 生物統計 / 回帰の発見
 - ピアソン (数学者): 生物統計から数理統計学 / 記述統計を大成

統計学の歴史

- 記述統計から推測統計へ
 - 20世紀: 推測統計の成立
 - ゴセット: ビールの醸造の研究から Student's t 分布を発見
 - フィッシャー: 農学での実験計画研究から推定 / 仮説検定
 - ネイマン / エゴン・ピアソン: 推測統計学の構築
 - 信頼区間・仮説検定などの理論的体系
 - ベイズ統計: 主観的な確率
 - 事前分布の導入 / 階層ベイズ
 - 21世紀: あらゆる分野で活躍
 - 経済・物理・疫学・生物・インターネット・マーケティング
 - コンピュータの活用



この講義で学ぶこと

- 本講義が目指すところ
 - 今後の研究・生活・社会/経済活動に不可欠な統計に関して、データを集計・分析し、理解する力を養う
- 研究のため
 - 実験、試験、調査などの結果を用いた実証研究を行う上でなくてはならないもの
- 生活のため
 - 生活に関わる様々な効果やリスクがデータとともに語られ、生活者としても統計に対するリテラシーが求められる
- 社会・経済活動のため
 - 情報技術の発展により日々膨大なデータが生成されるようになり、その有効活用が求められる

この講義で学ぶこと

- 本講義の概要
 - 統計の基本的な考え方を中心に講義
 - 厳密な数学的証明は避け、統計・統計学のエンドユーザーとして必要とされる直感的な理解を目指す
 - 一連の手順を解説することが中心
 - データの収集・チェック・集計・分析・結果の解釈
 - 適切な手法を選択することが極めて重要
 - そのためには少々の理論的な背景も知っておく必要
 - 二元分割表の独立性の検定と関連性の強さの推定
 - 統計ソフトウェアを使った自習
 - これらを正しく使い適切な結果を導けるようになることを目指す

この講義で学ぶこと

- 到達目標
 - 調査や実験・試験によるデータ収集の作法を理解する
 - データの種類や性質に応じたデータ確認と要約ができる
 - 二元分割表の独立性の検定と関連性の推定を行い、結果を解釈する
 - 仮説検定や推定の原理を理解する
 - 統計や統計学的知識を正しく使うための留意点と倫理を知る
 - 統計・統計学の応用について幅広く知り、今後の学習につなげる

平均値・最大値・最小値の計算

- あるクラスのある科目の得点リスト

学生番号	1	2	3	4	5	6	7	8	9	10
得点	3	4	5	9	2	9	7	9	7	8

- 平均点・最高点を求めよ

学生番号 <i>i</i>	1	2	3	4	5	6	7	8	9	10
得点 x_i	3	4	5	9	2	9	7	9	7	8
総和 $S_i = x_1 + \dots + x_i$	3	7	12	21	23	32	39	48	55	63

$$\text{平均点 } m = \frac{S_N}{N} = \frac{63}{10} = 6.3$$

学生番号	1	2	3	4	5	6	7	8	9	10
得点 x_i	3	4	5	9	2	9	7	9	7	8
最高点 $M_i = \max(x_1, \dots, x_i)$	3	4	5	9	9	9	9	9	9	9

$$\max(\max(a, b), c) = \max(a, \max(b, c)) = \max(a, b, c)$$

データの整列

- データを整列するには様々な方法がある

学生番号	1	2	3	4	5	6	7	8	9	10
得点	3	4	5	9	2	9	7	9	7	8
学生番号	1	2	3	4	5	6	7	8	9	10
得点	3	4	5	9	2	9	7	9	7	8
学生番号	1	2	3	5	4	6	7	8	9	10
得点	3	4	5	2	9	9	7	9	7	8
学生番号	1	2	5	3	4	6	7	8	9	10
得点	3	4	2	5	9	9	7	9	7	8
学生番号	1	5	2	3	4	6	7	8	9	10
得点	3	2	4	5	9	9	7	9	7	8
学生番号	5	1	2	3	4	6	7	8	9	10
得点	2	3	4	5	9	9	7	9	7	8

データの整列(続き)

学生番号	5	1	2	3	7	9	4	7	8	10
得点	2	3	4	5	7	7	9	9	9	8
学生番号	5	1	2	3	7	9	4	7	8	10
得点	2	3	4	5	7	7	9	9	9	8
学生番号	5	1	2	3	7	9	4	7	10	8
得点	2	3	4	5	7	7	9	9	8	9

学生番号	5	1	2	3	7	9	10	4	7	8
得点	2	3	4	5	7	7	8	9	9	9

分散と平均の関係

- 各データとの差の平方の総和を最小にする $x \rightarrow$ 平均値

- 最小化すべき関数

$$f(x) = (x_1 - x)^2 + (x_2 - x)^2 + \dots + (x_N - x)^2$$

- 導関数

$$f'(x) = -2(x_1 - x) - 2(x_2 - x) - \dots - 2(x_N - x) \\ = 2Nx - 2(x_1 + \dots + x_N)$$

$$f'(x) = 0 \text{ となる } x \text{ を求める } \quad x = \frac{x_1 + \dots + x_N}{N} = \bar{x}$$

「平均」は偏差の平方の総和を最小にする

中央値・最頻値の計算

- データを昇順 (あるいは降順に) 整列させる

学生番号	1	2	3	4	5	6	7	8	9	10
得点	3	4	5	9	2	9	7	9	7	8

学生番号	5	1	2	3	7	9	10	4	6	8
得点	2	3	4	5	7	7	8	9	9	9

- データ $x_1, \dots, x_N$ を整列した結果を $x'_1, \dots, x'_N$ と表すとき

$$\text{中央値: median}(x_1, \dots, x_N) = \begin{cases} x'_{(N+1)/2} & N \text{ が奇数のとき} \\ \frac{x'_{N/2} + x'_{N/2+1}}{2} & N \text{ が偶数のとき} \end{cases}$$

データの整列(続き)

学生番号	5	1	2	3	4	6	7	8	9	10
得点	2	3	4	5	9	9	7	9	7	8
学生番号	5	1	2	3	4	7	6	8	9	10
得点	2	3	4	5	9	7	9	9	7	8
学生番号	5	1	2	3	7	4	7	8	9	10
得点	2	3	4	5	7	9	9	9	7	8
学生番号	5	1	2	3	7	4	7	8	9	10
得点	2	3	4	5	7	9	9	9	7	8
学生番号	5	1	2	3	7	4	7	8	9	10
得点	2	3	4	5	7	9	9	9	7	8
学生番号	5	1	2	3	7	4	7	9	8	10
得点	2	3	4	5	7	9	9	7	9	8

分散・標準偏差・変動係数

- 分散 σ^2 : データの散らばりを表す

$$\sigma^2 = \frac{(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + \dots + (x_N - \bar{x})^2}{N}$$

- 標準偏差 σ : 分散の正の平方根

- 正規分布なら $\text{mean} \pm \sigma: 68\%$, $\pm 2\sigma: 95\%$, $\pm 3\sigma: 99.7\%$

- 変動係数 CV (coefficient of variation) $\frac{\sigma}{\bar{x}}$

- 相対標準偏差 (relative standard deviation: RSD) とも呼ばれる

- 平均値が異なる二つの集団のばらつきを比較するのに用いる

- 偏差値 T_i : x_i を平均値 50・標準偏差 10 となるようにスケールした値

$$T_i = \frac{10(x_i - \bar{x})}{\sigma} + 50$$

分散の計算

- 分散 σ^2 : データの散らばりを表す

- 定義に従うとまずは平均を計算する必要がある (データを二度走査)

$$\sigma^2 = \frac{(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + \dots + (x_N - \bar{x})^2}{N}$$

- 以下のように変形するとデータを一度走査するだけで計算可能

$$\begin{aligned} (x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + \dots + (x_N - \bar{x})^2 \\ = x_1^2 + x_2^2 + \dots + x_N^2 - 2\bar{x}(x_1 + x_2 + \dots + x_N) + N\bar{x}^2 \\ = x_1^2 + x_2^2 + \dots + x_N^2 - 2\bar{x}(N\bar{x}) + N\bar{x}^2 \\ = x_1^2 + x_2^2 + \dots + x_N^2 - N\bar{x}^2 \end{aligned}$$

したがって

$$\sigma^2 = \frac{1}{N} \sum_i x_i^2 - \left(\frac{1}{N} \sum_i x_i \right)^2$$

分散の計算

26

- データを一っぺつ加算していけば計算可能
- 総和 & 平方の総和にそれぞれ加算していく

学生番号 i	1	2	3	4	5	6	7	8	9	10
得点 x_i	3	4	5	9	2	9	7	9	7	8
得点の平方 x_i^2	9	16	25	81	4	81	49	81	49	64
総和 $E_i = x_1 + \dots + x_i$	3	7	12	21	23	32	39	48	55	63
平方の総和 $V_i = x_1^2 + \dots + x_i^2$	9	25	50	131	135	216	265	346	395	459

数値データの集計: 代表値で比較

27

- 代表値で比較するのが簡便
- 夜食を食べる人・食べない人の体重を比較
- ヒストグラムそのものではなく、平均と標準偏差を比較

表 夜食習慣別の平均体重

夜食	体重	回答数	平均 (kg)	標準偏差
する		24	58.50	4.93
しない		76	50.32	4.07
計		100	52.45	4.44

統計入門 第3回 質的変数(調査方法)

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
(情報学研究科併任)
林 和剛
hayashi.kazunori.4w@kyoto-u.ac.jp

調査の方法

28

- 正しい調査のためにはそれなりの「お作法」がある
- 課題調査の設定
 - 現状把握型
 - 仮説検証型
- 集計・解析方法
- 調査対象の設定 (標本の取り方)
- 調査時期
- 調査方法

調査課題の設定

29

- 調査目的
 - 現状把握型
 - 「大阪人の食生活」はどうなっているだろうか
 - 仮説検証型
 - 「『大阪人の食生活はかわっている』という仮説を検証したい
- 調査課題の設定と調査項目
 - 現状把握型
 - 調査課題: 「大阪人の食事メニューを調べる」
 - 調査項目: 「昼食メニューは次のうちどれですか？」
 - 仮説検証型
 - 調査課題: 「大阪の人はたこ焼きが好き人が多い」
 - 調査項目: 「大阪人ですか?」「たこ焼き器を持っていますか?」

調査対象の設定: 母集団と標本

31

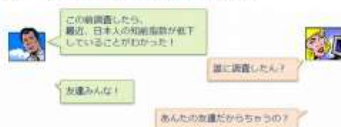
- 全体を代表する標本をとることが大切



調査対象の設定: 標本の取り方

32

- 有意抽出法は選択バイアスを生む
- 便宜的方法
 - 年齢・性別・地域・所属など、なにかの属性に基づいて調査対象の選定を行う方法
 - 主観に基づくため標本の偏りに客観的な評価ができない
- 標本の偏り: 選択バイアス (selection bias) を生む



調査対象の設定: 標本の取り方

33

- 多くの場合で無作為抽出法が望ましい
- 単純無作為抽出
 - 母集団 (電話帳、住民台帳、その他名簿など) に適し番号をつけ、くじ・乱数などによって必要な数を選ぶ
- 系統抽出: 母集団の名簿に番号をつけ一定間隔で標本を選ぶ
- 多段階抽出: 母集団からの抽出を何段階かに分けて行う方法
 - 全国調査時に、最初に調査する都道府県を無作為に選び、都道府県の中から市町村を選び、最後に市町村の中から人を選ぶ
- 層化抽出: 既知の母集団の状況 (男女比等) に合わせて抽出
- 集落抽出: あらかじめ母集団をグループ分けしておいて、グループを無作為に選び、選ばれたグループ全員を調査

「統計の入門」#1 イントロダクション (2/2)

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

統計学の分類

■ 統計には「記述統計」と「推測統計」がある

- 記述統計
 - 全数調査を前提
 - データを整理・視覚化することで理解
- 推測統計
 - サンプル調査を前提
 - 部分から全体を知る
 - 仮説が正しいかを判断する
 - 過去から未来を予測する
- 両者は確率論でつながる
 - 標本抽出時のゆらぎを考える

医学・医療における「統計」のはじまり

医学・医療における「統計」のはじまり

フロレンス・ナイチンゲール

生誕 | 1820年5月12日
没年 | トスカーナ大公国フィレンツェ
1905年10月13日 (85歳没)
イギリス・ロンドン
著名な業績 | 近代看護教育の父
統計学発展者
専門 | 看護学、統計学
業績記号 | 花柳病

ジョン・スノウ

生誕 | 1813年3月15日
没年 | 1858年3月18日 (45歳没)
イギリス・ロンドン
著名な業績 | 近代疫学教育の父
統計学発展者
専門 | 疫学、統計学
業績記号 | 花柳病

医学・医療における「統計」のはじまり

■ 兵士の死因（クリミア戦争 | 1854年）

元祖「見える化」

医学・医療における「統計」のはじまり

■ コレラ

- 激しい下痢からの脱水で死に至る
- 別名「虎狼病」
- コレラ菌を病原体とする経口感染症
- 現代日本では稀
- 現代医療で制菌は容易

医学・医療における「統計」のはじまり

■ 1854年のロンドン

- 特定の井戸周辺にコレラ多発

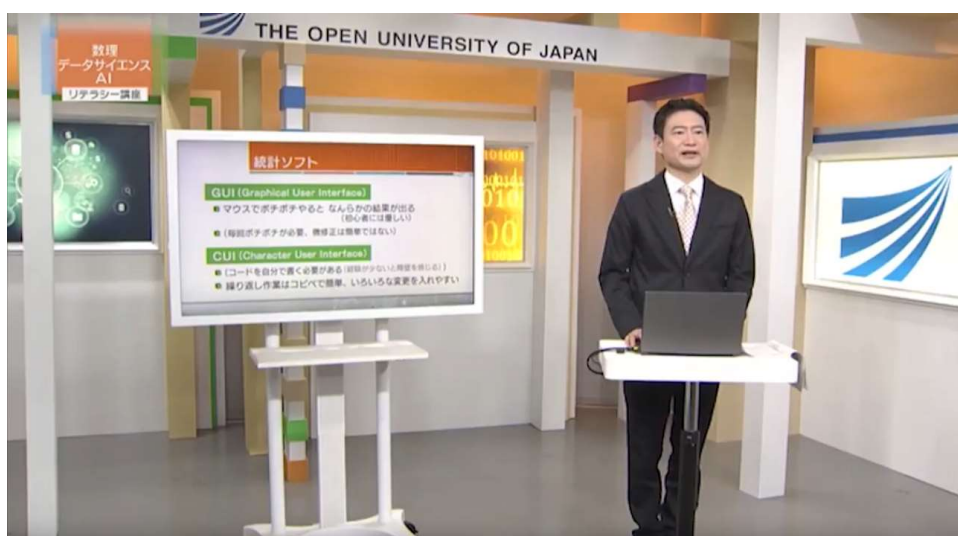
医学・医療における「統計」のはじまり

■ 水道水とコレラ

同一の区域でも、水道供給会社によって「データを集めること」の重要性

コレラ発生率に大きな違いがあった！

放送大学放送の様子



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 の 裏 側 に

健 康 維 持 の 裏 側 に、

と も だ ち の 始 ま り の 裏 側 に、



京都大学全学共通教育でのデータサイエンス教育の特色

1 全学共通	2 最先端を提供	3 Society5.0に即応
文理を問わず、どの学部・研究科であっても学べるように全学共通科目として提供* *履修・卒業要件は、各学部・研究科の指示に従ってください。	● 課外学習としてデータサイエンススクールなどの実施 ● 多岐にわたる分野の研究所・企業等と連携	全ての学生がSociety5.0時代に必要となる数理的思考力とデータ分析・活用能力を体系的に身に付けることができる

 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の公式アカウント公式ツイッター[[@cirads2020](#)]でも情報発信中。

Follow Me!

京都大学国際高等教育院全学共通科目

自然科学科目群 データ科学分野 履修案内

令和3年度 全学共通科目 自然科学科目群 データ科学分野 科目一覧

科目名	内容
統計入門	● 全学部、全回生向け（医学部医学科は1回生必須科目） ● 統計学の基本的な考え方を理解します ● 統計解析ソフトを用いた演習を自習形式で取り入れています
統計と人工知能	● 全学部、全回生向けとして「統計入門」で扱い切れなかった話題を扱います ● 人工知能・機械学習の基本的な考え方、手法を理解し、統計学がどのような形でその基礎を形作っているか学びます（分散分析、回帰分析、因果推論、深層学習、画像処理など）
数理統計	● 主として2回生 理系学部学生向け ● 確率論基礎ならびに微分積分学、線形代数を既知とします ● 確率論に基礎を置く推定・検定を重視した伝統的な数理統計の基礎を理解します
データ分析基礎	● 大規模なデータ解析を行うための基礎的な技術を得得します ● 厳密な数学的証明は必要最低限、直感的な理解を深めます ● 回帰分析、主成分分析、Excelと統計ソフトRを使った実習を行います
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とします - 演習例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 - 演習例2：医学・医療データとExcel、R言語を用いた統計処理
数理・データ科学のための数学入門 I, II	● データ科学を学ぶために必要な最低限の知識の習得を目標にしています ● 数学的な理論の完全な体系ではなく、データ科学への必要性を説いてから内容を厳選します ● 文系学部学生向けに高校での数学IIIの知識を仮定しません

令和3年度 大学院共通・横断科目群 データ科学イノベーション教育研究センター 担当科目一覧

科目名	科目群	内容
データ科学概観	大学院共通科目群	● 様々な学術領域で研究を開始するために必要なデータ科学手法について学びます ● 実習を通じて、Excel、R言語、Python等を用いてデータ分析に必要な基礎的な処理ができるようになります
データ科学展望 I, II	大学院横断科目群	● 実際に大規模なデータ解析を行うための手法に触れます ● 幅広い選択ができるように、異なる専門分野を背景に持つ教員、異なる講義内容を用意しています
データ科学展望 III, IV	大学院横断科目群	● データ科学イノベーション教育研究センターが提供するData Science School（下記）において、学内外の専門家による講義・演習によってデータ科学的な手法を体験しながら実践的に学びます
データ科学：理論から実用へ I, II データ科学：理論から実用へ演習 I, II	大学院共通科目群	● 統計学・数理科学分野の発展的内容であるベイズ統計・マルコフ連鎖モンテカルロ法（MCMC）、データ同化・数値シミュレーションなどの基本的な考え方や実装方法を基礎から解説します ● 演習では統計解析ソフトを使った実習を行います

大学院横断教育統計・情報・データ科学系科目群には、上述の他にも各研究科が多くの科目を提供しています。詳細は国際高等教育院大学院横断教育科目群のHPを参照してください。

Data Science School

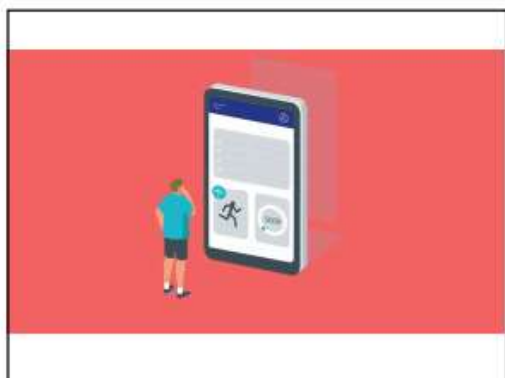
データサイエンススクール

「数理及びデータサイエンスの係る教育強化」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ことができ、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セミナーを企画・開催しています。詳しくはセンターHPをご覧ください。



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
Center for Innovative Research and Education in Data Science
Institute for Liberal Arts and Sciences, Kyoto University
〒606-8315 京都市左京区吉田道徳町69 道徳館202, 301, 302号室
Konoekian 202, 301, and 302, Yoshida Konoekian-cho 69, Sakyo, Kyoto 606-8315, Japan
Tel. 075-753-9691 URL <http://ids.kyoto-u.ac.jp/>





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

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。

お客様のデータサイエンスは、京華の大学卒業後随分と学ぶことが出来ます

あ の 宅 配 の 裏 側 に も

最 高 の 買 い 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**「データサイエンス(AI)」**の基礎的な必要能力を全ての国民が育み、あゆむ分野で人材の活躍

主要施策

- 国内産の人材を育成・伸ばす環境整備**
 - ① 最先端の独自研究と海外研修の機会を提供
 - ② 知識をAIで活用し、解決する課題への課題解決型AI人才能育成
- AIGC活用への啓蒙**
 - ① AIと専門技能のタッグアンドアップ促進
 - ② AIで地場産業の発展につなげる人材育成（厚労省）
- 認定制度・資格の創設**
 - ① 大学や専門学校で実習プログラムを政府が認定する制度確立
 - ② 民間企業、IT大手（ゼネラル）の奨励、中小企業向け支援
- 学習環境の強化**
 - ① 文科省主導で子ども向け検定試験（MOE）活用
 - ② 高校におけるAIの必修化と実践教育の充実
- 中高校に迫る教育格差の解消**
 - ① 多様な人材の確保（国産技術に人員上、IT人材は外国に乏しい）

実施時期【2025年】

- ① 大学卒業生 2,000人/年
② 大学院生 200人/年
③ 職業訓練生 100人/年度/7年
- ④ 認定制度 20万人/年
⑤ 民間企業 50万人/年
⑥ 文科省主導 100万人/年
⑦ 民間企業 300万人/年
⑧ 文科省主導 100万人/年

すべての大学卒業生が修得しているべき素養

<https://www.moe.go.jp/kambo/saze/brand/attach/pdf/as-1-5.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(数理科学)の融合分野

理論より証拠
証拠(データ)から理論

情報学

統計学

数理科学

基礎
情報
数学
分析
確率
統計
金融
政策
決定
知能
大数
データ
科学
生体
科学
人工
知能
情報
科学

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

6

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 → **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
$$P_{\theta}(A) = \frac{P_{\theta}(B) P(A)}{P(B)}$$

$$c_n = \frac{1}{n} \sum_{i=1}^n f(i) \cos \pi i \theta, (n = 0, 1, 2, 3, \dots)$$

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

- 惑星は一定の軌道に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

天動説の場合は、真円軌道に小さな周回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

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

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

流行ピーク、8割がインフルエンザの外発

感染性

地区	人口	死亡	感染性
東部	1440	31	2.1%
西部	1000	400	40%
計	2440	431	17.6%

検査結果が陽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
$$P_{\theta}(A) = \frac{P_{\theta}(B) P(A)}{P(B)}$$

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 (自然系)	情報群
リテラシーレベル (全同学向け) 情報・統計に関する一般常識・基本知識	• 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	• 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等)
専門基礎レベル (主に2年生以上向け) データから分布などの情報を抽出する力	• 統計と人工知能 • データ分析演習 I・II • データ分析基礎	• 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.shuglin.go.jp/internet/index.nsf/html/images/gjldou002.jpg/\\$File/gjldou002.jpg](http://www.shuglin.go.jp/internet/index.nsf/html/images/gjldou002.jpg/$File/gjldou002.jpg)

*4 鎌谷 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:Shuglin-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)