全学データサイエンス 2単位必修化: 初年度から2年度に向けて

> 筑波大学 佐久間 淳 (共通科目「情報」推進室)

数理・データサイエンス教育強化の流れ

…データに内在する本質的構造を見極め…データサイエンスを活用して…有用なシステム構築につなげる能力が求められる。

文部科学省 数理及びデータサイエンス教育の強化に関する懇談会 (H28.12)

文系も含めて全ての大学生が一般教養として数理・データサイエンスを履修できるよう、標準的な カリキュラムや教材の作成・普及を進める。 未来投資戦略 2018 閣議決定 (H30.6.15)

文系理系を問わず、必要とされる数理・データサイエンスの基礎的素養を持つ・・・・・人材育成を推進。 文部科学省 大臣提出資料 (H30.9.28)

- 筑波大学ではH29年度からDS必修化の実施体制検討に着手
- H30年度に体制構築・教材整備
- R1年度から全学1年次2100名にDS必修化

筑波大学におけるDS実施方針

- DSは理工系だけのものではない
 - リテラシーを扱う1年次共通科目として設置
 - 2 単位全学必修 (2100名, 150分/週*10回)
- DSはデータを実際に触って学ぶべき
 - 全クラス計算機室で実施
 - 75分の座学+75分の計算機演習(主にエクセル)
 - 1クラス 学生約50名/教員 1, TA 1 (演習指導補助)
 - 日本語クラス50+英語クラス1

筑波大学のDS実施体制

- 実施体制
 - 推進室教員10名 (DS教育を推進する情報工学域の常勤教員)
 - 常勤教員17名 (情報工学域他所属教員)
 - 非常勤教員21名 (大学8名、研究所4名、企業8名、他1名)
- 推進室としての活動
 - 担当教員に標準教材を作成・提供
 - スライド、小テスト、演習課題、10回分
 - 専門性の高い内容はビデオ講義を提供(後述)
 - 各分野DS活用、ビッグデータ、IoT/CPS、機械学習、AI
 - FD機会の提供
 - 学生アンケート、教育効果測定(後述)、教員懇談会

DSシラバス

授業計画	講義概要
第1週~第5週	社会におけるデータサイエンス:
	データに基づく課題解決,データサイエンスの様々な事例,データに関する法規と倫理
	データの収集:
	データサイエンスの基本プロセス,データの種類, データの収集,データの前処理,
	データの再利用性
	データの管理:
	データ管理の意義と目的,データ収集項目の設計,情報構造と表現の分離,
	高度なデータ管理とビッグデータ
第 6 週~第 10 週	データの可視化:
	可視化の意義と目的,視覚的表現の選び方
	データの分析:
	離散変数の理解と統計,量的変数の理解と統計,因果と相関,時系列データ,ネットワークデータ,
	高度なデータ分析と人工知能

- R1開始のため、開始時点ではモデルカリキュラムを参考にはできなかった
- ・結果的にはモデルカリキュラム(数学基礎、計算 基礎、モデリングと評価以外の項目)に準拠

教育支援システムによる資料提供

筑波大学

University of Tsukuba

※自己登録許可期間中

2020-08-04 (Tue)

佐久間 淳 | 設定 | ログアウト





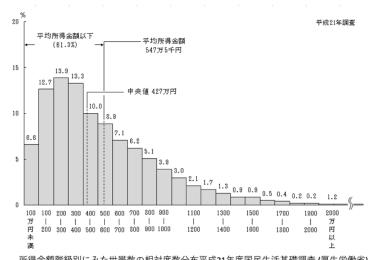
記述統計の比較

- 複数の記述統計を比較するには、データは同一の条件下 で収集される必要がある
- A高校の学生のセンター入試「数学」の平均点は2018年度 68.3点、2019年度は72.6点だった。2019年度の学生は2018 年度の学生よりも成績が良い?
 - 問題の難易度が変化したかもしれない
 - ・ 2018年度は多くの成績の良い学生がセンター入試を必要としな い大学を受験したかもしれない
- 複数の統計値を安易に比較すると間違った結論を導くこ とがある

QUIZ 7-3

- 1. 日本の所得の最頻値はいくらか?
- 2. 日本において、なぜ平均値は中央値、最頻値より大きいの
- 3. 日本とアメリカで代表的な世帯所得を比較する場合、平均 値、中央値、最頻値のどれが適切か?

ヒストグラムを読む: 所得分布 (日本)



所得金額階級別にみた世帯数の相対度数分布平成21年度国民生活基礎調査 (厚生労働省)

https://www.mhlw.go.jp/toukei/saikin/hw/k-tyosa/k-tyosa09/2-2.html

Lecturer only

Quiz 7-3 **解答例**

- 1. 200-300万円
- 2. 高所得側に外れ値(非常に所得が高い世帯)が存在する。そのよ うな世帯が平均値を高所得側に引きあげる(このような理由から、 アメリカのヒストグラムでは平均所得は記載されていないと考えられ
- 3. Q2の考察から、この分布では平均を代表値とみなすのは適当で はない。また、最頻値については階級がことなり比較できない。よっ て以下の中央値を代表値とみなすのが適当である
 - 中央值: 427万円, US 53,657 USD

教育支援システムによる小テスト

DS第八回 QUIZ 3			
課題に関する説明			
受付期間	2019-12-02 14:50~2019-12-02 14:59		
選択肢のシャッフル	シャッフルしない		
ポートフォリオ	回答を学生のポートフォリオに追加		
採点結果と正解の公開	受付終了時に採点結果と正解を公開		

※採点シートに表示される問題番号を赤の太字で表示しています(例: 1.1)。

以下の二つの変数x,yは高い相関が観察されたとする。x,yの因果関係を否定する場合に、x,yについて適切であると考えられる関係を選びなさい。交絡因子が存在する場合には、何が交絡因子となっているかを書きなさい。

離婚数が減少すると、一人当たりマーガリン消費量が減少する (単一選択)

1.1

- 2. 交絡因子の存在
- 3. 逆の因果
- 4. 偶然の一致

交絡因子が存在する場合には、何が交絡因子となっているか

1.2

足のサイズが大きい小学生は、数学の成績が良い (単一選択)

1.3

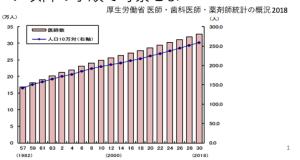
- 1. 相関関係
- 2. 交絡因子の存在
- 3. 逆の因果
- 4. 偶然の一致

教材事例(演習パート)

座学で理解した知識を、データを使った演習で実践

演習9-2 日本の医師不足問題

- ・日本の医師数は増加している(下図)
- 医師不足が問題になっているのはなぜか?
- 次ページ以降の手順で考察せよ



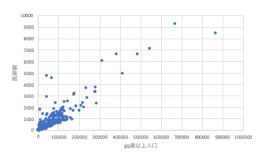
-

Lecturer only

医師数と高齢者数の関係

高齢者の多いところは医師数も多い 仮説と違う?

.... 人口が交絡因子になっている



3

問題意識: 地域による医師の偏在

- 「医師にかかることの多い高齢者数の多い地域 に医師が少ないことが問題ではないか?」と考 えたとする
- そこで、「高齢者数の多い地域に医師が少ない」という仮説が正しいかどうかを全国の市区 町村のデータを用いて検証する
- まず以下から市区町村別データをダウンロード せよ (演習 9-2-1)
 - https://www.nstac.go.jp/SSDSE/

2

人口あたりの医師数と 高齢者数の関係

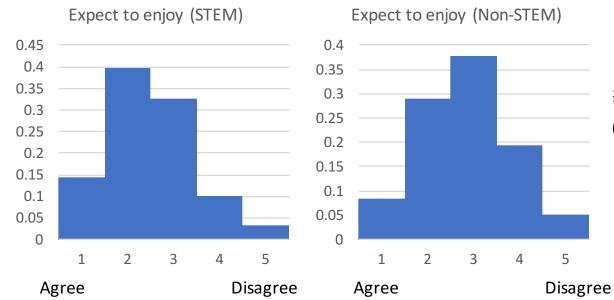
- 医師数と高齢者数の関係には「人口」という交絡 因子が存在する。
 - つまり、人口が多ければ医師も高齢者も多くなる
- そこで人口の影響を除くため1人あたりの高齢者数 および1人あたりの医師数の関係を調べることを考 える
- 人口一人あたりの高齢者数および1人あたりの医師 数を求めよ (演習 9-2-3)
 - その際、人口0人の市町村のデータを除くこと
- さらにそれらの関係を散布図でプロットせよ (演習 9-2-4)

4

2

DS全学必修化の課題

- 筑波大学がカバーする教育分野は広い
 - 人文・文化、社会・国際、人間、生命環境、理工、 情報、医学、体育、芸術
- •特に非理工系の18歳の学生は、理工系の学生に 比べて、DSへの興味が薄い



教育効果測定(2019) Q. (DSの)授業が楽しみだ。

学ぶ動機を高める2つの仕掛け

- 1. 自分が学ぶ分野とDSの橋渡し (R1年度から)
 - 各教育組織のデータを活用した研究を手掛ける教員のビデオ講義をカリキュラムに組み込む
 - 演習教材として、様々な分野のデータを準備
- 2. 自分たちのデータを自分で解析(R2年度予定)
 - 自分が解析対象となることで、DSを自分ごととして捉える (研究倫理委員会申請中)

学ぶ動機を高める(1):専門家によるビデオ講義

- ・初回講義でアラカルト形式で3本選択
- ・視聴後、小テスト

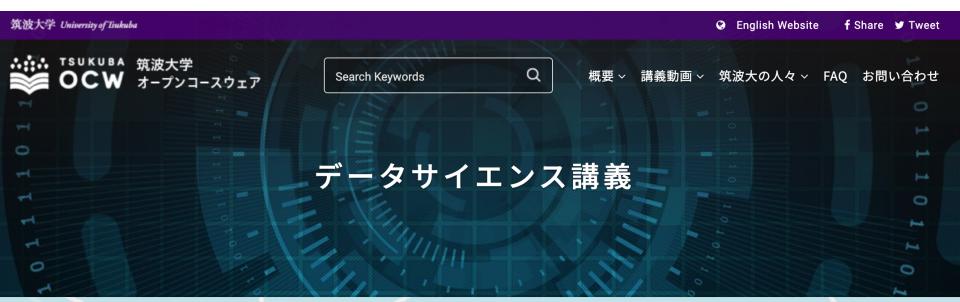
R1年度

教員所属	タイトル
人文社会系	デジタル・ヒューマニティーズ—人 文学と情報学 の接点が導く新たな知 識の世界
数理物質系	仮説検定入門
図書館情報 メディ ア系	ヒューマンインタラクション
体育系	現代サッカーボールの空力特性
医学医療系	生命科学とデータサイエンス
医学医療系	臨床医学・社会医学とデータサイエンスー ヘルス サービスリサーチの視点から

R2年度追加予定

教員所属	タイトル
人文社会系	海洋における放射能汚染の推移に 関 するオープンデータ
生命環境系	人工 衛星データによる森林伐採の安 全な飲み水へのアクセ スの影響分析
システム情 報系	AI による人 権侵害と公平性の確保
システム情 報系	シミュレーションにおる 津波の予測 と防災
体育系	サッカー画像解析と 戦略立案

OCWにて全講義公開中 https://ocw.tsukuba.ac.jp/data-science/



↑ Home / データサイエンス講義

概要

筑波大学では、平成31年度10月より、情報教育に関する共通科目として「データサイエンス」を開講しました。「データサイエンス」の授業では、データサイエンスの基礎的概念を理解し、コンピュータを利用した基礎的なデータ分析技術を学ぶことを目的としています。データの収集、データの管理、データの可視化、データの分析を通じて、データの理解と活用の手法を実践的に習得し、先端的なデータサイエンスの事例に触れ、社会におけるデータの具体的な活用について理解することを目指します。本サイトでは、各学問分野におけるデータの収集、管理および活用について学ぶために使用するビデオ講義を公開しています。

医学医療系 田宮菜々子教授

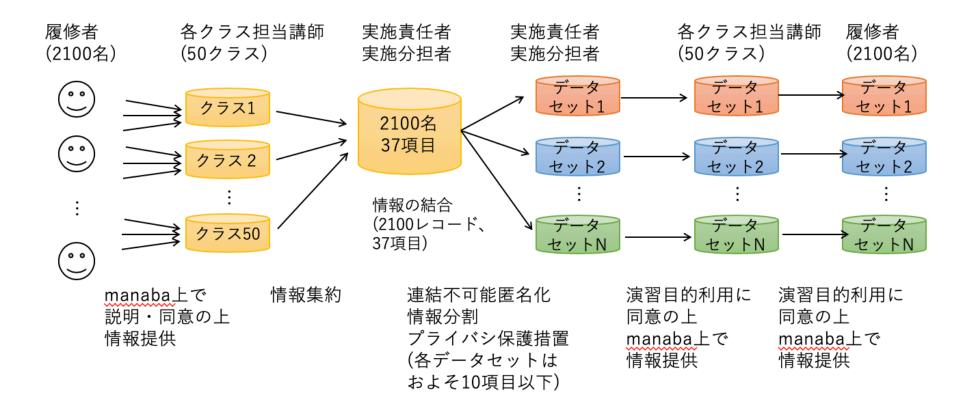
臨床医学・社会医学とデータサイエンスーヘルス サービスリサーチの視点から



システム情報系 馬場雪乃 准教授 人工知能と機械学習



学ぶ動機を高める(2): 自分たちのデータを自分で解析(R2より)



第二週目に学生からデータ収集 プライバシ保護処理等の後、第六週目に講師にデータ提供

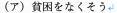
学生に対する質問項目例

2. 2015年に国際連合にて193ヶ国の首脳により署名された持続可能な開発目標 (SDG)、またはグローバル目標はご存知ですか。

(ア)はい

(イ)いいえ 」

3. SDG の 17 個のグローバル目標のうち、どのグローバル目標が、あなたとあなた の家族にとって目前に迫っている課題ですか? (6つ選択してください。) 🗸

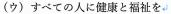


・あらゆる場所で、あらゆる形態の貧困に終止符を打つ↓

(イ) 飢餓をゼロに↓



・飢餓に終止符を打ち、食料の安定確保と栄養状態の改善を達成するとと もに、持続可能な農業を推進する↓



・あらゆる年齢のすべての人の健康的な生活を確保し、福祉を推進する↓



(エ) 質の高い教育をみんなに↓



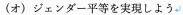
(a)

Ů×††∗Ť

2 fine

3 すべての人に 健康と福祉を

・すべての人に包摂的かつ公平で質の高い教育を提供し、生涯学習の機会 を促進する。



・ジェンダーの平等を達成し、すべての女性と女児のエンパワーメントを 図る4

(国際連合広報センター MY WORLD 2030)

<将来の仕事に関する質問>↓

7. 職業選択の重視点

仕事を選ぶ際に、どのようなことを重視しますか。この中からいくつでも選んで ください。

- (ア)収入↓
- (イ)労働時間。
- (ウ)通勤の便₄
- (エ)仕事内容
- (オ)職場の雰囲気↓
- (カ)仕事の社会的意義。
- (キ)事業や雇用の安定性↓
- (ク)将来性』
- (ケ)専門的な知識や技能が生かせること
- (コ)能力を高める機会があること↓
- (サ)自分を生かすこと
- (シ)自分の好きなことや趣味を生かせること
- (ス)その他。
- (セ)わからない。

(内閣府 我が国と諸外国の若者の意識に関する調査(平成30年度)から) し

教育効果測定

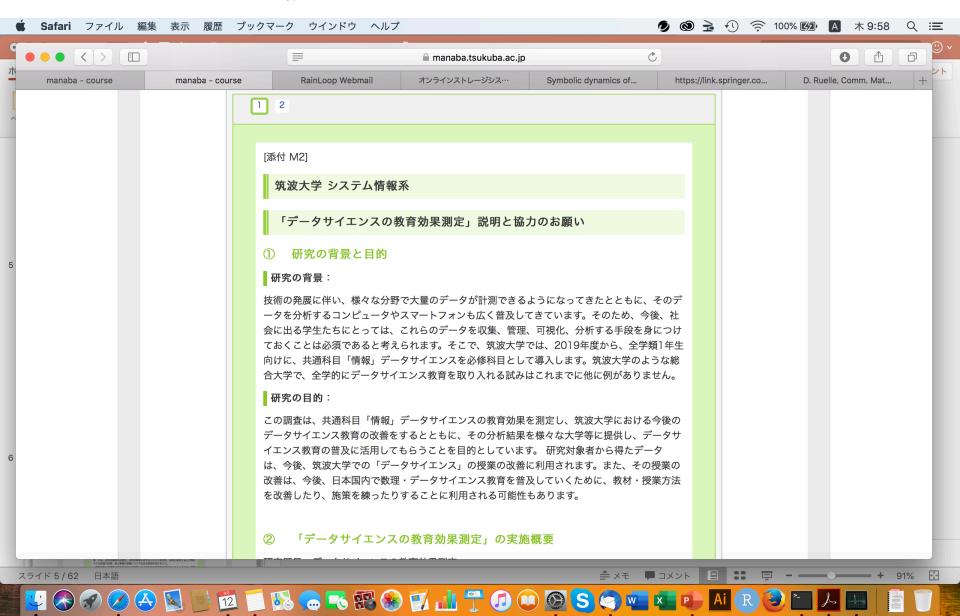
• 趣旨

- 全履修者の教育効果を測定し、カリキュラムの継続的改善に役立てる
- カリキュラム設計に有益な情報を学外に発信

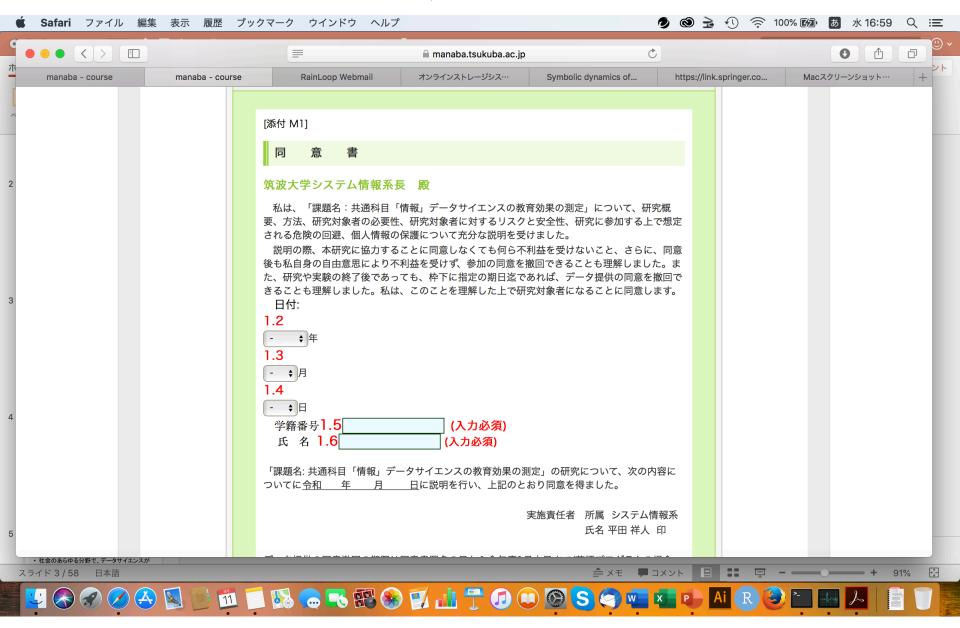
• 調査内容

- 対象: DSを履修する1年次全学生
- 時期: 2019-2021年
- 第一週: 趣旨説明、同意書、意欲と用語理解 (主観調査)
- 第九週: 用語理解 (主観調査)、DS理解度(客観調査)
- 学生アンケート、講師アンケートと連結調査 (倫理申請中)

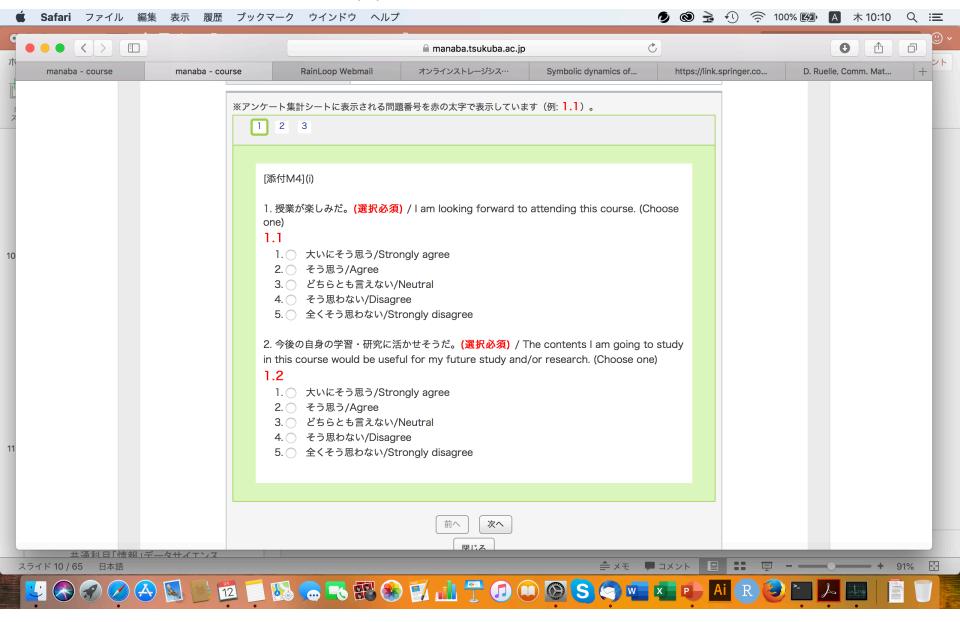
教育効果測定 趣旨の説明



教育効果測定 同意書 2019年度、調査に同意し全データを提供した学生は6割強



教育効果測定 動機に関する質問



教育効果測定 用語の理解 (主観評価)

※アンケート集計シートに表示される問題番号を赤の太字で表示しています(例: 1.1)。1 2 3 4 5 6 7 8 9
以下の用語のうち、言葉の意味が説明できるものにチェックを入れなさい。 1.3 1. □ 母集団 2. □ データ欠損 3. □ データベース 4. □ ヒストグラム 5. □ 分散 6. □ 交絡因子 7. □ 幾何平均
前へ 次へ 閉じる

教育効果測定結果の分析: 学生の理解につながる要因はなにか?

- 「Q5 授業の内容をどの程度理解しましたか?」に強く 関連する項目 (線形回帰)
- p<0.0001
 - 「Q3. 授業を通じて、この科目に関連する分野への興味や関心が高まりましたか。」
 - →興味を高める事が重要
- p<0.05
 - 「授業が楽しかった」「授業で習ったことは、今後、どの程度 役に立つと思いますか。 |
 - →将来の学習・研究やキャリアに結びつけることが重要
- 関連なし
 - 授業開始前の「期待」や「役に立つ」という思いは、理解度に あまり関係しない

未解決課題・今後の展望

- 問題解決能力の養成
 - データ収集項目・データ解析課題の立案と解析
- 学生の背景知識に合わせた教材の調整

いずれもスケールしにくい実施内容で、 実施は今後の課題

お知らせ

- R1年度教育効果測定の詳しい解析については... 第二回首都圏ブロックWS 9/17 10:00-12:00 (オンライン)
- 筑波大学DS教育の詳細については… オペレーションズ・リサーチ誌第65巻第11号 「筑波大学における全学必修のデータサイエンス教育」 to appear.
- オープンコースウェア
- https://ocw.tsukuba.ac.jp/data-science/ どなたでも閲覧可能です。ご利用される場合は info-literacy-inquiry@u.tsukuba.ac.jpまでご連絡を