筑波大学における教材の内容

今倉暁（筑波大学）

もくじ

- データサイエンス授業概要
  - 講義全体の概要
  - 学習の範囲

- 教材内容について
  - 講義
  - 演習
  - ビデオ講義

データサイエンス授業概要

全体の概要

- 目的
  - データサイエンスの基礎を、座学およびエクセルを用いた演習を通して、しっかり学ぶこと

- そのため、
  - 実社会への応用については深くは取り扱わない
    - 各分野に対応した応用については、ビデオ講義を用意
  - 演習であつかうデータは、エクセルで扱える範囲に限る
    - データには人工データ、オープンデータの他、自分たちのデータも利用する
    - プログラミングに興味がある人は、各学類の講義を（開講されていれば）受講を勧める

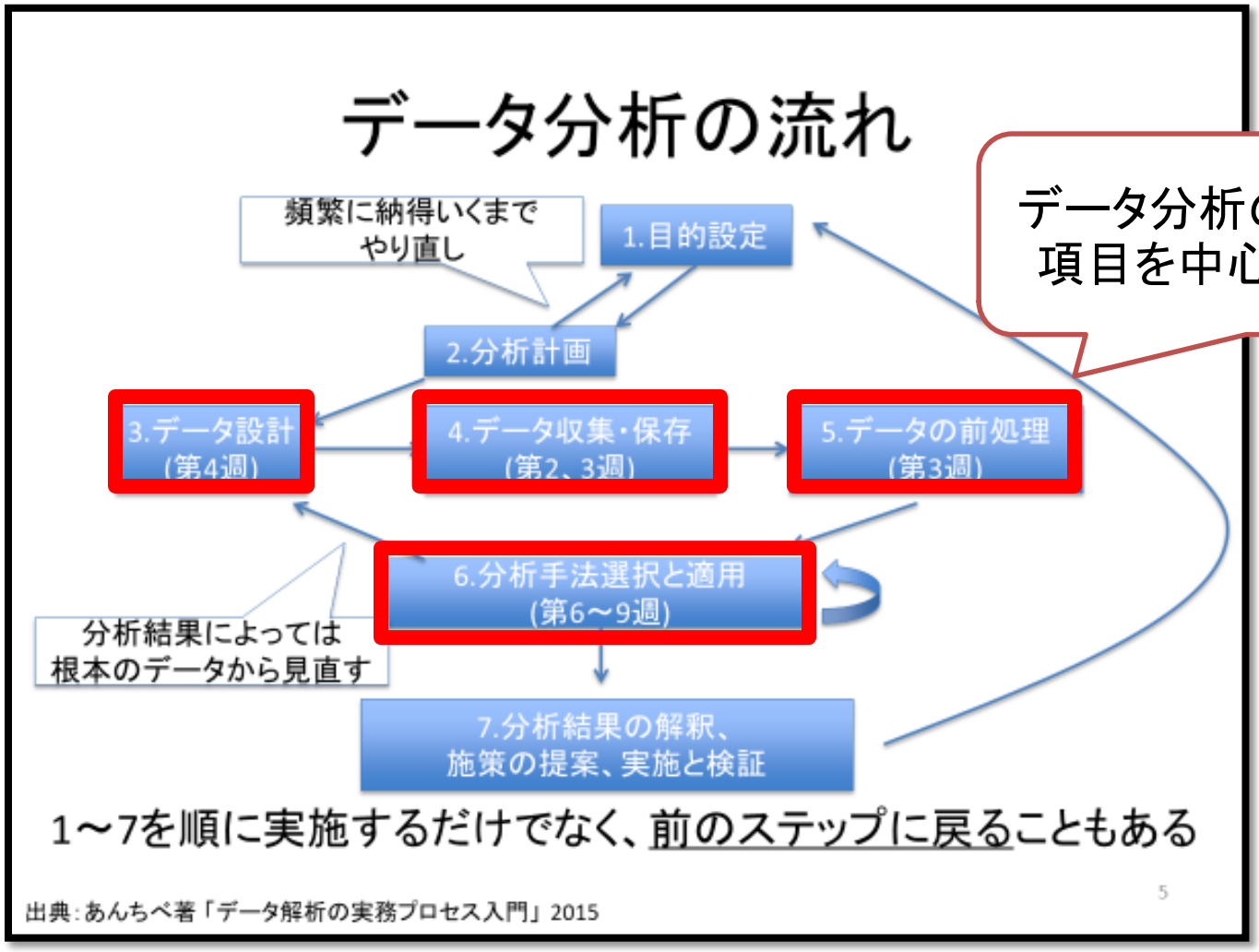
全体の概要

- 授業形態（150分/週 x 10週）
  - 講義、演習、ビデオ講義
  - 小テスト、レポート

- 全学必修であるための工夫
  - 講義資料および演習内容については、下記の3種類を用意
    - 簡易版：数理に馴染みのない文系クラスを想定
    - 標準版：数理に馴染みのある文系クラスを想定
    - 発展版：理系クラスを想定
  - 多様な学生の興味と動機の向上を目的とした各分野に対応したビデオ講義を用意（詳細は後述）
  - 自分たちのデータを自分たちで解析する実践的な演習課題（詳細は後述）

学習の範囲

- データサイエンスの基礎 （授業資料の抜粋）



データ分析の流れで重要な
項目を中心に各週で学ぶ

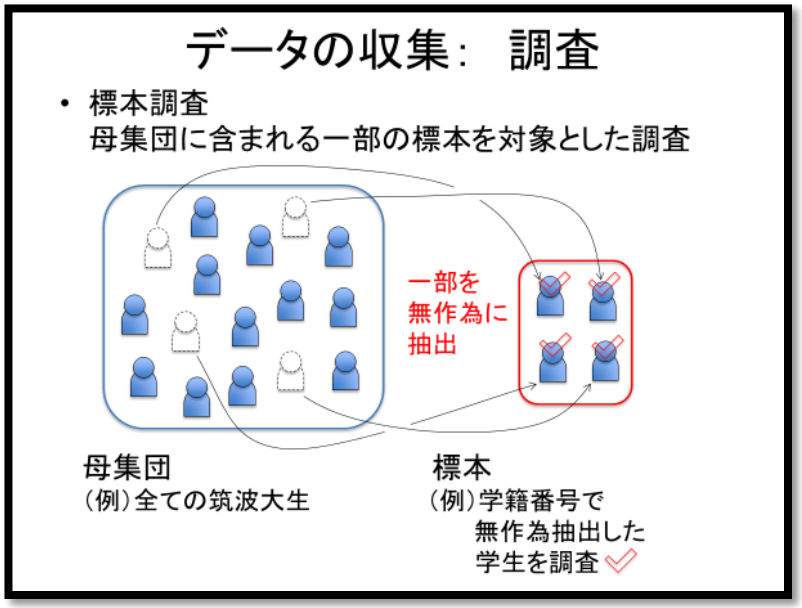
教材内容について

講義内容・キーワード

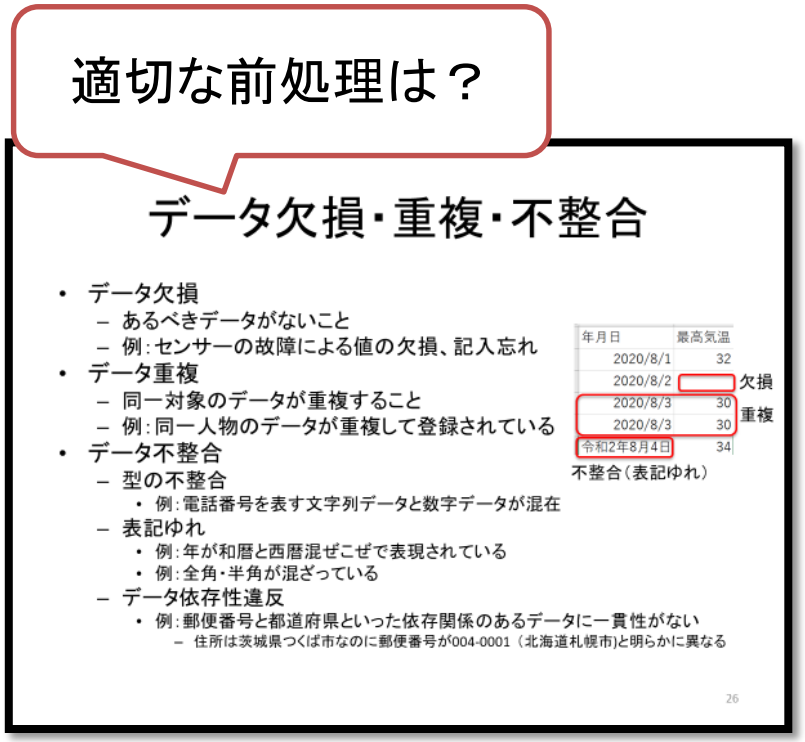
週	講義内容	キーワード
1	データサイエンス概論	データサイエンスの概要
		データサイエンス導入ビデオ講義
2	データの種類と収集1	データの種類、研究倫理、データの収集、エクセル基本操作
3	データの種類と収集2	データサイエンスの基本プロセス、データの前処理
		データの再利用性、エクセル応用操作（統計）
4	データの種類と収集3	データ管理の意義と目的、データ収集項目の設定
		情報構造と表現の分離、エクセル応用操作（データ操作）
5	高度なデータ管理 達成度評価（中間）	ビッグデータビデオ講義
6	データの可視化と質的データの解析	可視化の意義と目的、視覚的表現の選び方、データ可視化演習
		1つの質的データ(度数分布表)、2つの質的データ(クロス集計表)
7	1つの量的データの解析	1つの量的データの理解、ヒストグラム、記述統計、平均値の違い
		統計検定
8	2つの量的データや時系列データの解析	2つの量的データの理解，相関と相関係数，時系列データ，（回帰直線）
9	実際のデータの分析	様々なデータの分析、ネットワークデータ、相関と因果
10	高度なデータ分析 達成度評価（期末）	人工知能ビデオ講義

データの種類と収集

- 内容
  - データの種類と尺度水準、収集方法と調査方法、研究倫理、エクセルの基本操作
  - データサイエンスの基本プロセス、特にデータの収集法と前処理法



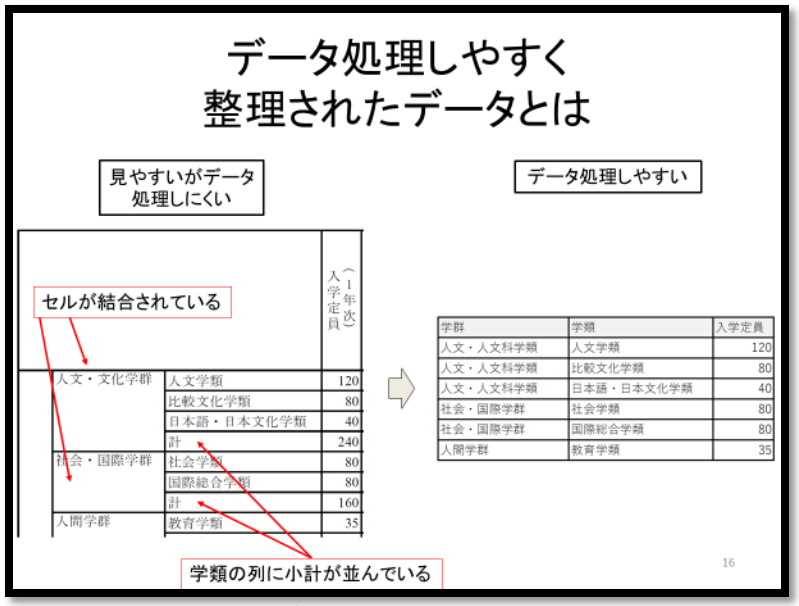
データ収集・調査方法



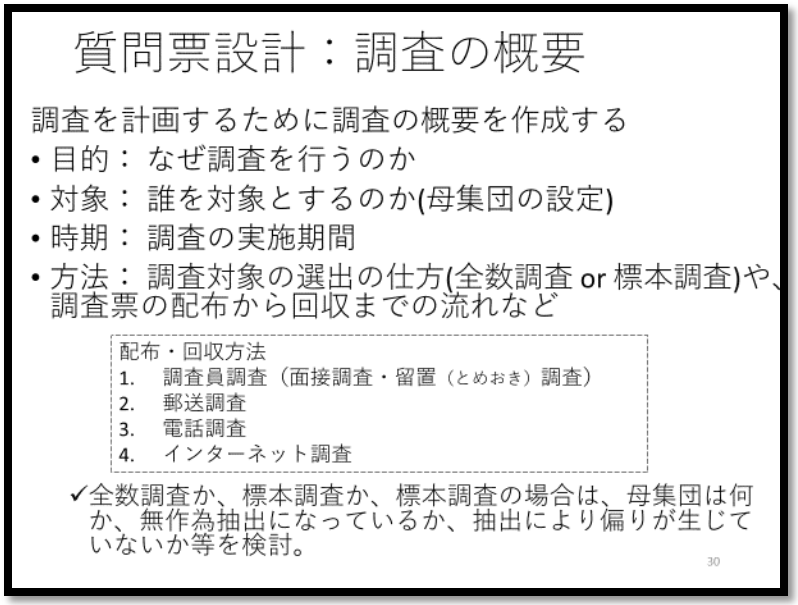
データ前処理

データの種類と収集

- 内容
  - データの収集と管理
    - 文書構造（データ構造）と視覚表現（ビュー）を分離することの意義
    - 質問票の設計



データの管理



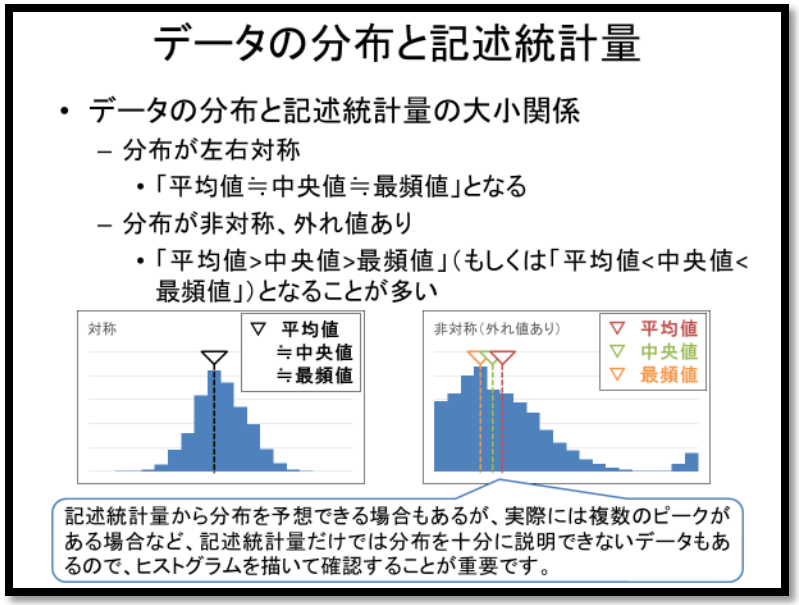
質問表の設計

講義内容・キーワード

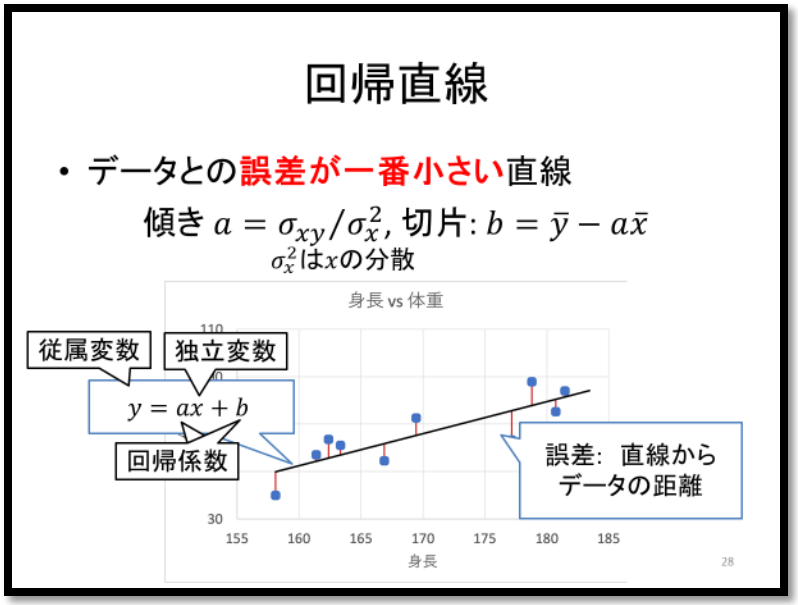
週	講義内容	キーワード
1	データサイエンス概論	データサイエンスの概要
		データサイエンス導入ビデオ講義
2	データの種類と収集1	データの種類、研究倫理、データの収集、エクセル基本操作
3	データの種類と収集2	データサイエンスの基本プロセス、データの前処理
		データの再利用性、エクセル応用操作（統計）
4	データの種類と収集3	データ管理の意義と目的、データ収集項目の設定
		情報構造と表現の分離、エクセル応用操作（データ操作）
5	高度なデータ管理 達成度評価（中間）	ビッグデータビデオ講義
6	データの可視化と質的データの解析	可視化の意義と目的、視覚的表現の選び方、データ可視化演習
		1つの質的データ(度数分布表)、2つの質的データ(クロス集計表)
7	1つの量的データの解析	1つの量的データの理解、ヒストグラム、記述統計、平均値の違い
		統計検定
8	2つの量的データや時系列データの解析	2つの量的データの理解，相関と相関係数，時系列データ，（回帰直線）
9	実際のデータの分析	様々なデータの分析、ネットワークデータ、相関と因果
10	高度なデータ分析 達成度評価（期末）	人工知能ビデオ講義

データの分析

- 内容
  - 1つの量的データの代表値やヒストグラムの読み方、異なる母集団の平均値の違い(Z検定)
  - 2つの量的データの可視化や関係性の解析方法、時系列データの可視化や解析方法



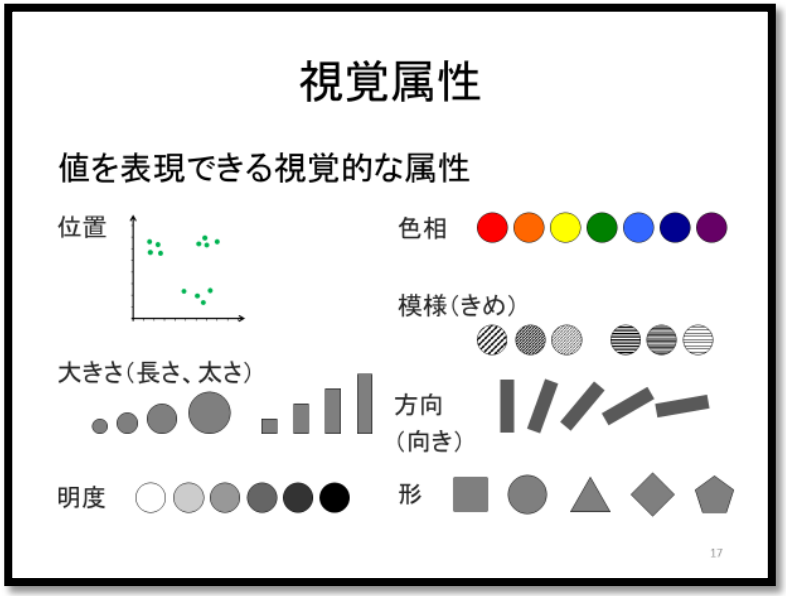
データの分布



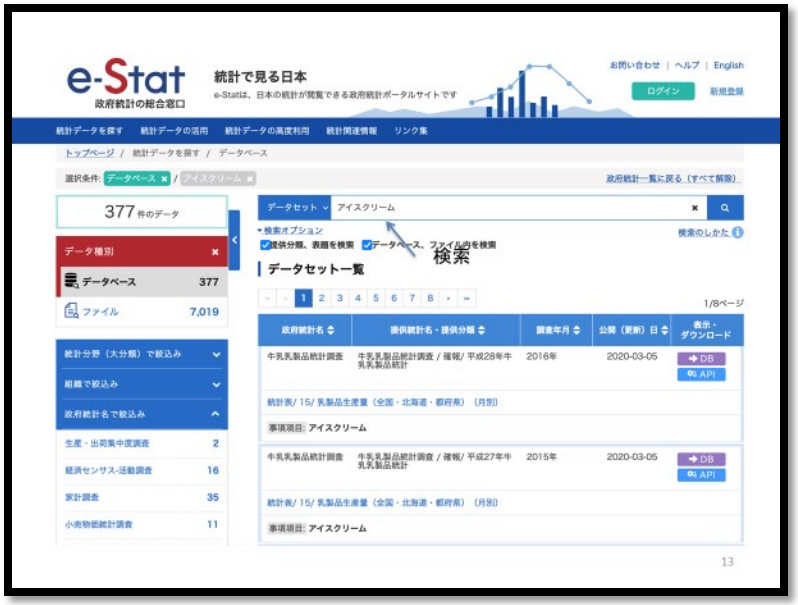
2つのデータの関係性

データの分析

- 内容
  - 可視化：データに適したグラフ（チャート）の選び方や描き方、不適切なグラフとその理由
  - 実際のデータ解析の流れの学習
  - Excel では扱えないような複雑なデータ解析の事例についても紹介



可視化で使用する視覚属性

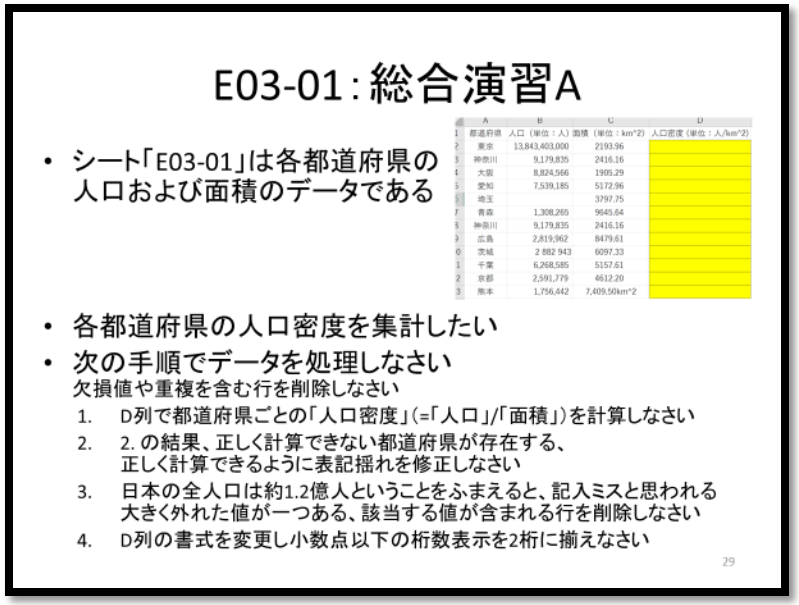


e-Statからのデータ収集

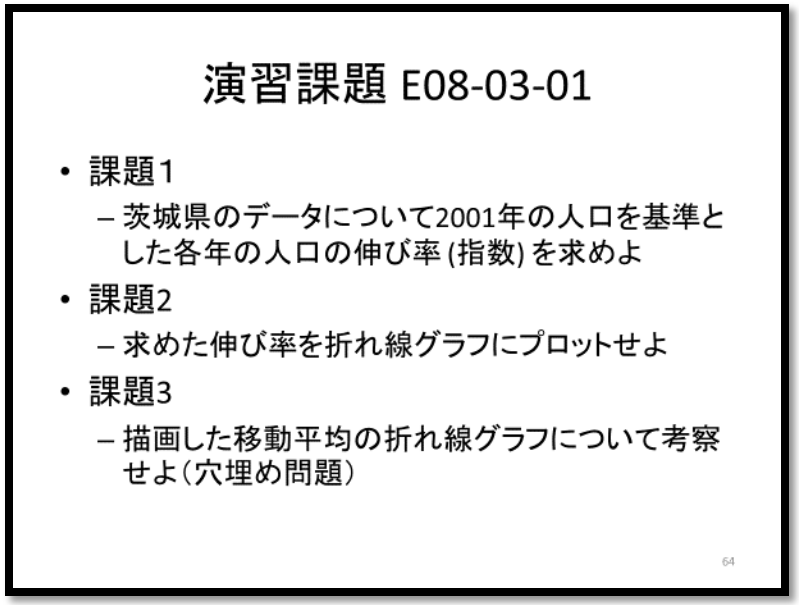
演習

- 概要
  - エクセルを使用して、各講義内容の演習に取り組む
    - エクセルの基本操作、データの前処理
    - 可視化、データの入手、データの分析

    など



データ前処理



データの分析

演習

- 使用するデータ
  - 人工データ
  - オープンデータ
    - タイタニック、カリフォルニアハウジング
    - 気象庁、e-Stat
  - 自分たちで集めたデータ
    - アンケートデータ

自分たちのデータを自分で解析

学生

各クラス担当教員指導のもと
全学生(参加任意)からの
データ収集(第二週)

共通科目「情報」推進室

データ集約
匿名化&プライバシ保護
課題例作成
データ提供(第五週)

各クラス担当教員

授業にて自分達の
データをデータ解析
(第六週以降)

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。）

(ア) 貧困をなくそう
・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

(ウ) すべての人に健康と福祉を
・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
・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

(オ) ジェンダー平等を実現しよう
・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

(国際連合広報センター MY WORLD2030)

<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
(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)

問題意識(SGD)、キャリア意識、
およびITリテラシーに関する質問項目から構成

講義内容・キーワード

週	講義内容	キーワード
1	データサイエンス概論	データサイエンスの概要
		データサイエンス導入ビデオ講義
2	データの種類と収集1	データの種類、研究倫理、データの収集、エクセル基本操作
3	データの種類と収集2	データサイエンスの基本プロセス、データの前処理
		データの再利用性、エクセル応用操作（統計）
4	データの種類と収集3	データ管理の意義と目的、データ収集項目の設定
		情報構造と表現の分離、エクセル応用操作（データ操作）
5	高度なデータ管理 達成度評価（中間）	ビッグデータビデオ講義
6	データの可視化と質的データの解析	可視化の意義と目的、視覚的表現の選び方、データ可視化演習
		1つの質的データ(度数分布表)、2つの質的データ(クロス集計表)
7	1つの量的データの解析	1つの量的データの理解、ヒストグラム、記述統計、平均値の違い
		統計検定
8	2つの量的データや時系列データの解析	2つの量的データの理解，相関と相関係数，時系列データ，（回帰直線）
9	実際のデータの分析	様々なデータの分析、ネットワークデータ、相関と因果
10	高度なデータ分析 達成度評価（期末）	人工知能ビデオ講義

ビデオ講義

- 授業内容の客観評価を高めるには学生の興味と動機向上が重要
- 様々なバックグラウンドを持つ学生の特性を考慮した23本のビデオ講義
- オープンコースウェアとして誰でも視聴可能(https://ocw.tsukuba.ac.jp/)

筑波大学オープンコースウェア
UNIVERSITY OF TSUKUBA OPENCOURSEWARE

筑波大学 10学群（学生の背景）

- 人文・文化学群
- 社会・国際学群
- 人間学群
- 生命環境学群
- 理工学群
- 情報学群
- 医学群
- 体育専門学群
- 芸術専門学群
- 総合学域群

ビデオ講義一例（データ駆動型社会における津波高即時予測より）

ビデオ教材で専門に近い動画が用意されている学類の学生は、ビデオ講義で学習動機が向上
（授業アンケート結果より）

ビデオ講義

各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

全学群の
学生が視聴

データサイエンスの導入

- ヒューマンインタラクション
- 臨床医学・社会医学とデータサイエンス
- 現代サッカーボールの空力特性
- 生命科学とデータサイエンス
- デジタル・ヒューマニティーズ
- 人工知能における倫理的・法的・社会的問題
- データ駆動型社会における津波高即時予想
- サッカーの上達にデータを生かす
- データサイエンスと社会科学
- セシウム137全球データベースおよび
  環境放射能データの検索と公開サイト

高度なデータの管理と活用

- ビックデータとIoT/CPS
- 人工知能と機械学習

データサイエンスの発展

- 仮説検定入門

まとめ

- データサイエンス講義の目的
  - データサイエンスの基礎を、**座学**および**エクセル**を用いた演習を通して、しっかり学ぶこと

- 教材内容
  - データサイエンスの基礎
    - **データ収集、前処理、可視化、分析**
  - 学習意欲の向上への取り組み
    - **各種分野でのデータ解析の応用に関するビデオ講義**
    - **自分たちのデータを演習で利用**