

統計学入門 補助資料 ～ガイドンス～

2022年度1学期: 月曜2限
担当教員: 石垣 司

1

授業の概要

- 本授業の内容: 大学以降の統計学の入門的内容
 - 高校数学(I, A, II, B)の知識を利用
 - 数学IIIや大学数学の知識があると理解が促進される
 - 経済経営数学入門や数学科目の同時履修を勧める
 - 記述統計学と推測統計学, 確率変数と確率分布, 統計的推測, 統計的仮説検定を理解する
- 授業の進め方
 - 講義と演習
 - 数理統計の授業は積み上げ型 ⇒ 復習を意識して欲しい

2

授業の日程

- 2022年7月25日までに全15回の授業を実施予定

- | | |
|---------------------------|----------------------------|
| 1. 4月11日(オンライン) | 9. 6月06日(G1:対面, G2:オンライン) |
| 2. 4月18日(G1:対面, G2:オンライン) | 10. 6月13日(G1:オンライン, G2:対面) |
| 3. 4月25日(G1:オンライン, G2:対面) | 11. 6月20日(G1:対面, G2:オンライン) |
| 4. 5月02日(オンライン) | 12. 6月27日(G1:オンライン, G2:対面) |
| 5. 5月09日(G1:対面, G2:オンライン) | 13. 7月04日(G1:対面, G2:オンライン) |
| 6. 5月16日(G1:オンライン, G2:対面) | 14. 7月11日(G1:オンライン, G2:対面) |
| 7. 5月23日(G1:対面, G2:オンライン) | 15. 7月25日(オンライン) |
| 8. 5月30日(G1:オンライン, G2:対面) | |

– 学籍番号下1桁が奇数の学生 ⇒ グループ1(G1)

– 学籍番号下1桁が偶数の学生 ⇒ グループ2(G2)

※ 休講があった場合もそれ以降の対面 or オンラインのグループはこの予定通りで実行

※ 授業実施形態は今後のCovid-19感染状況で変更することもある

3

授業の資料・連絡先

- Google Classroomに授業動画などをアップロード
 - 私のHPにも講義資料のみアップロード
 - http://www2.econ.tohoku.ac.jp/~isgk/lec_material/basic_stat2022.html
- Google Classroomコード:「yje35o7」
- 教員の連絡先
 - e-mail: isgk “at mark” tohoku.ac.jp
 - 履修登録などに関する質問などは随時受付
 - 数式の説明などはメールでは対応できないので不可
 - 最低限のマナーとして, 氏名と学籍番号を記載してメールを送信。その記載のないメールには返信しない

4

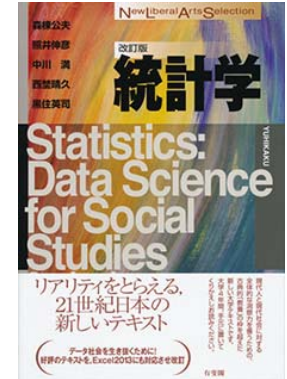
成績の付け方

- **成績の付け方: 演習問題レポート100%**
 - 毎回の講義で演習問題を出題
 - Googleフォームを利用してその回答を提出
 - Google Classroom内のレポート提出フォームを利用。メールや印刷物での提出は認めない
 - 毎回の「回答の方法」のファイルを参照すること
- **レポート提出締切: 授業の日の次週の午前9時**
 - 月曜2限の締切 ⇒ 次週月曜の午前9時
 - 締切後の提出は一切認めない
 - 成績確定後の“泣きつき/お願いメール”は一切認めない
 - ※履修の確認は必ず行うように

5

教科書・参考書

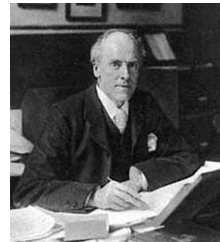
- **教科書: 無し (Web上で補助資料配布)**
 - もしも、資料の誤記を見つけたら是非教えてください
- **参考書**
 - 何でもよいので必要であれば、自分に合った参考書を探してください
 - 例えば、「森棟, 照井, 中川, 西埜, 黒住, 統計学, 有斐閣, 2015」など



6

統計学って?

- **統計学は科学の文法**
 - The Grammar of Science (Karl Pearson 1892)
 - 実証分析では必須の学問
 - 経済学, 経営学, 工学, 理学, 医学, 生物学, 心理学, 社会学, 教育学, 文学, 政治学など
- **統計学とは?**
 - 集団現象を観察し分析する方法を研究する学問(大辞林)
 - 確率論を基盤にして, 集団全体の性質を一部の標本を調べることによって推定するための処理・分析方法について研究する学問(デジタル大辞泉)
 - 本講義での設定
数学とコンピュータの力を借りて, 社会や学術上の問題を解決するためにデータを活用するための方法の学問



Karl Pearson 1857-1936

統計の起源～①国の実態をとらえるための統計

- **古代: 国家統治のための統計**(例:ピラミッド建設)
 - Statisticsの語源 ラテン語「Status」: 国家・状態
 - センサス(Census: 古代ローマで行われた人口調査)
- **16世紀以降の欧州: 国家間の勢力比較**
 - ドイツ: 国勢学
 - 国家の基礎事項(人口, 土地面積, 貿易規模など)を数量化。官庁統計
 - フランス: 統計局の設置(ナポレオン 1801)
 - 18世紀以降, 丁抹, 米, 蘭, 英などで初の近代的センサス
- **国勢調査(センサス)は現在でも実施**
 - 日本: 5年に1回(全国民)



統計の起源～②大量の事象をとらえるための統計

- 政治算術学派(17世紀イギリス)
 - 政治的解剖(社会構造を解剖し国政へ活用)
 - 大量観測により法則や因果関係の発見を重視
 - 数と量と尺度によって社会を対象とした議論を展開
- ベストの予測(John Grant)
 - 死亡統計表(生命表)(1662)の分析
 - ロンドンの人口の見積もり(200万人⇒約38万人)
- 保険料算出(Edmond Halley)
 - 当時の保険料はギャンブル要素強
 - 死亡統計表から合理的な保険料を算出
 - 生命保険事業の基礎



William Petty 1632-87
(参考書第1章表紙)

※総務省統計局補助教材より <http://www.stat.go.jp/teacher/c2epi1.htm>

統計の起源～③確率的事象をとらえるための統計

- ギャンブルの研究から確率論が誕生
 - 賭博(サイコロやトランプなど)の確率の仕組みを数学的に研究
- 16世紀
 - G. Cardano 標本空間の基本的概念(サイコロトランプの研究から)
 - G. Galileo 小論「サイコロゲームについての考察」
- 17世紀
 - B. Pascal & P. Fermat 確率論の基礎
- 18世紀
 - T. Bayes, J-L. Lagrange, P-S. Laplace等 確率論の大成
 - A. de Moivre 年金論
 - D. Bernoulli 天然痘の罹病&死亡率の計算
 - L. Euler 抽出調査から全体を推計する方法



※総務省統計局補助教材より <http://www.stat.go.jp/teacher/c2epi1.htm>

近代統計学(19世紀以降)

- 3つの流れを「統計」として整理
 - 社会統計に確率論を導入(A. Quetelet「近代統計学の父」)
 - 社会現象・自然現象を数量的にとらえる
- 記述統計学(K. Pearsonが大成)
 - データの平均・分散, 可視化などから分布を議論し, 対象の傾向や性質を把握
 - ヒストグラム, 標準偏差, 相関係数など
- 推測統計学(R. Fisherが体系化)
 - 部分の標本から全体の構造を推定
 - 仮説検定, 最尤推定, 実験計画など



Adolphe Quetelet 1796-1874



Karl Pearson 1857-1936



Ronald Fisher 1890-1962

この講義で扱う統計学

すでに始まっている近未来社会

- Cyber-Physical Systemによる超スマート社会
- Cyber-Physical System
 - 現実世界の観測データから, 仮想空間で情報・知識を創出し, 現実世界での価値を提供するシステム



内閣府HPより

- Society 5.0 ではデータが“資源”
 - AI技術, ビッグデータ活用があらゆる産業で浸透

AIブームの本質は“データ”

ビッグデータ時代の到来

- 1990年代 情報通信技術の発展
- 2000年代 情報通信技術の爆発的普及
- 2010年代 ビッグデータ・AIブーム, AI技術の発展
- 2020年代 AI技術の普及

人工知能ブームの本質 ⇒ データ + データ学習技術

現在の第3次AIブーム



将棋
プロ棋士(2013) 名人(2017)
囲碁
元世界ランク1位(2016) 現役世界ランク1位(2017)



画像認識
人間の精度を超える(2015)



自動翻訳
https://pocketalk.jp/



完全自動走行への期待
無人走行バス
地域を支える新しい交通手段、高齢者の移動もより便利に。

AI・データ人材の不足

世界的な規模で需要増

LinkedInのレポート

(IEEE Spectrum, 31 Aug, 2018)



- The biggest shortage of data science experts is in New York City (34,032), followed by the San Francisco Bay Area (31,798), and Los Angeles (12,251).
- But the biggest chasm in Silicon Valley is between demand for and supply of people who can get out from behind the keyboard and demonstrate oral communication skills (a shortage of 100,666).

経団連～政府への要望

(採用と大学教育の未来に関する産学協議会、中間とりまとめ共同提言2019年4月22日)

2. AI、数理統計、データサイエンス人材育成に向けた措置

- AI、数理・データサイエンス学部の新設を政策的に推進すべき
〔例:23区の定員規制について上記学部の新設は例外扱いとすること〕
- 統計学を教えられる専門人材の育成

※1大学設置基準、3規制措置、4地方創生事業に関して。分野に関する要望は2のみ

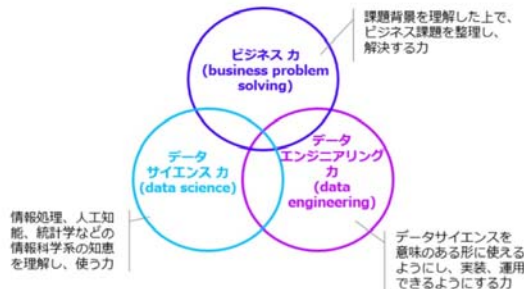
※2019年6月19日「統合イノベーション戦略」を閣議決定。2025年までにAI人材25万人育成を目指す

理学部や工学部の仕事でしょ？

必要とされる経済・経営がわかるデータ科学人材

- データ科学には数学・コンピューターの力が必要
- しかし、それら“のみ”では不十分

データサイエンティストに求められるスキルセット



データサイエンティスト協会HPより抜粋

Advantage 1 ▶
5年で修士号が取得可能なプログラム

データ科学
日本の経済・経営
高齢化社会の地域経済政策
高度グローバル人材コース

東北大学経済学部では、学部4年 + 修士1年 の5年で修士号取得が可能な4つの一貫プログラムを用意しました

Advantage 2 ▶
ビジネス力を備えたデータサイエンティストの養成

● 現代社会では、実業・産界の知識をもとに備えた人材が求められています
● データを分析してビジネスに役立てるアナリスト、コンサルタント、マーケターなどの活躍の場が広がっています

経済・経営学の知識を修得
経済学、経営学、会計学、統計学、経済史

Society 5.0で活躍できる人材

修士課程
統計・計量分析特別演習
データ科学特別演習

3・4年次
中級計量経済学特講
上級計量経済学特講
Applied Data Sciences

2年次
計量分析
ファイナンス
オペレーション・リサーチ
応用実証分析

1年次
経済経営数学基礎
統計学入門

理系の知識を活用

データ科学人材の待遇

米ではデータサイエンティストや統計家は好待遇

米CareerCast.comの分析結果(220職種)

- 米国労働省労働統計局のデータを加工。環境、収入、将来展望、ストレスに関する数十種の得点の総合順位。収入の数値は中間値

	2017	2018	2019	2021	
1位	Statistician	Genetic Counselor	Data Scientist	Data Scientist	\$98,230
2位	Medical Services Manager	Mathematician	Statistician	Genetic Counselor	\$85,700
3位	Operations Research Analyst	University Professor	University Professor	Statistician	\$92,270
4位	Information Security Analyst	Occupational Therapist	Occupational Therapist	Medical Services Manager	\$104,280
5位	Data Scientist	Statistician	Genetic Counselor	Mathematician	\$110,860
6位	University Professor	Medical Services Manager	Medical Services Manager	University Professor	\$80,790
7位	Mathematician	Data Scientist	Information Security Analyst	Operations Research Analyst	\$86,200
8位	Software Engineer	Information Security Analyst	Mathematician	Information Security Analyst	\$99,730
9位	Occupational Therapist	Operations Research Analyst	Operations Research Analyst	Actuary	\$111,030
10位	Speech Pathologist	Actuary	Actuary	Software Engineer	\$110,140

※最高収入(2019)は「Surgeon (\$255,110)」だが、順位は109位