

第3回統計エキスパート育成に向けたワークショップ
日時 2023年8月31日 10:00-12:00
方法 zoomによるリモート開催

データ関連人材育成 プログラムの取組と 統計エキスパート人材育成

狩野 裕

大阪大学・MMDSセンター長

統計エキスパート人材育成プログラム
運営委員会委員(参画機関)



統計エキスパート人材育成プログラム

(Program for Training Experts in Statistical Sciences, TeSS仮称)

- 目的
 - DXやAIに必要なスキルを有する大学統計教員（統計エキスパート人材）を育成
- 対象
 - PD・若手教員
 - 広く諸科学分野を対象とする
 - 数理・データ・AIを専攻する人材に限らない
- ミッション
 - 統計学を用いた他分野との融合領域の研究振興を図る
 - 育成された大学統計教員が大学等で核となり、統計学や融合領域に係る教育や活動の普及・展開を図る
 - 統計エキスパート人材育成エコシステムの確立
 - 30名以上の大学統計教員を育成（研修生）
 - その30名が500名の統計エキスパートを育成する（孫研修生）

AI戦略2019 における目標

トップレベル
100

エキスパート

2,000

応用基礎

250,000

(理系学生 +
文系学生の1/3)

500,000

(大学・高専全員)

リテラシー

データ関連人材育成プログラム

- データサイエンス等のスキルを習得させる研修プログラムを開発・実施し、キャリア開発を支援
- アカデミアにおける多様な分野の研究者や、コンソーシアムに参加する企業等で活躍する高度人材を育成
- H29年に4コンソーシアム、その後1コンソーシアムを採択

数理・データサイエンス・AI教育の全国展開

- 全国的なモデルカリキュラム・教材の開発、教育に活用可能な社会の実課題・実データの収集・整備等
- 6大学を拠点校として設置。更に30大学を協力校・特定分野協力校、61大学を連携校として設置（令和2年8月31日現在）
- 全国の国公立大学等への普及・展開
- 私立大学の取組への支援強化

課題：統計の教育・研究の中核人材の不足

統計エキスパート人材 育成プロジェクト（新規）

統計学の教育・研究のエキスパート人材を育成

- 人材育成プログラムと共同研究により統計学のエキスパートに育成
- 育成された人材が大学等で核となり、統計学や融合領域に係る教育・普及を行う好循環体制を構築
- 統計学を用いた他分野との融合領域の研究振興を図る
- 大学・大学共同利用機関法人等のコンソーシアムを公募により支援

平成29・30年度公募 「データ関連人材育成プログラム」

- 目的

- 高度データ関連人材の卵を発掘し、関連する知識・スキルの習得, キャリア開発を行う
- 高度データ関連人材の活躍を促進し、データ利活用社会のエコシステム構築へ貢献

- 対象

- 研究活動を通じて高度なデータの扱いに親しんだ博士課程学生(DC)や博士号取得者(PD)等を対象(博士人材)
 - 数理・データ・AIを専攻する人材に限らない
 - 修士課程学生も対象に

- 方法

- **企業**や大学等が人材の発掘・育成・活躍促進を目的としたコンソーシアムを形成
- **インターンシップ**や**PBL**(課題解決型学習)等の実践的な研修プログラムを開発・実施

- ミッション

- 高度データ関連人材を広く供給する

選定された5つのプログラム

- 東京医科歯科大学
 - ビッグデータ医療・AI創薬コンソーシアム
- 電気通信大学
 - データアントレプレナーフェロープログラム
- 大阪大学
 - データ関連人材育成関西地区コンソーシアム
 - 全国ネットワーク
- 早稲田大学
 - 高度データ関連人材育成プログラム
- 北海道大学
 - 次世代スマートインフラ管理人材育成コンソーシアム

各コンソーシアムのプログラム概要 (＋全国ネットワーク)

代表機関	コンソーシアム	参画機関等	プログラム概要・科目
北海道大学	次世代スマートインフラ管理 人材育成コンソーシアム https://www.mdsc.hokudai.ac.jp/	東京大学大学院情報学環	本事業のプログラムは、以下2つの人材育成プログラムにより構成する。 1. 共同研究型データ関連PBLプログラム 2. 実践的デザイン力養成プログラム 1. 共同研究型データ関連PBLプログラム ・産学連携型PBLへの参加に必須となる「秘密情報や発明・知的財産の扱いについての知識を養成する研修」 ・本事業が育成を目指す人材像に必須となるソフトスキルを養成する研修 2. 実践的デザイン力養成プログラム ・実際の企業における課題に取り組み、問題の所在の明確化から解決方策のデザインまでを主導的に実施する。
早稲田大学	高度データ関連 人材育成プログラム https://www.waseda.jp/inst/cds/education/d-data	お茶の水女子大学 上智大学 湘南工科大学	「高度な専門性」×「データ科学」を兼ね備えた博士人材の育成を目的 様々な専門・バックボーンを持つ博士学生に対し、オンデマンドコンテンツをはじめ、自身の専門分野の研究やキャリアデザインにデータサイエンスを活用するための教育プログラムの提供や学生支援を実施 基礎研修プログラム: 【基礎研修科目】統計シリーズ、データ科学シリーズ、数学・ICTシリーズ 【セミナー】ハンスオンセミナー: アンケート分析、因子分析、自然言語処理、画像処理 データ科学入門フォローアップセミナー等 実践研修プログラム: データサイエンスコンペティション、データ科学インターンシップ、オンライン研究発表会、海外実践研修プログラム、データ科学シンポジウム
東京医科歯科大学	医療・創薬データサイエ ンスコンソーシアム http://md-dsc.com/	東北大学 慶應義塾大学 東京理科大学	入門編(導入編): データサイエンス入門、データサイエンスのための基礎数学、Linux入門、プログラミング入門 基礎編: データサイエンス基礎Ⅰ、データサイエンス基礎Ⅱ、ゲノム医学基礎、マネジメント特論、知的財産特論、生命倫理 プログラミング実習Ⅰ(R)、Ⅱ(Python)、医療統計解析実習(基礎編)、SQL入門。可視化&GUI実装実習、深層学習の基礎技術テキストマイニング、RIによるデータ前処理実習、他 応用編: ゲノム統計学概論、疾患データ科学特論、モバイルヘルス特論、AI創薬特論、システム医学概論、疾患データ科学特論、オミックス解析統計モデリング、ネットワーク解析、ブロックチェーン入門、GWAS解析、医療統計解析実習、人工知能・機械学習概論、他 実践編: 研修(企業人材コース・博士人材コース対象) 企業インターンシップ(博士人材コース対象) ワークショップ(共通)
電気通信大学	データアントレプレナー コンソーシアム https://de.uec.ac.jp/	津田塾大学 麻布大学 芝浦工業大学 東京農工大学 青山学院大学	データサイエンティスト素養とビジネス開拓マインドを身に付け実践的課題解決を体感し、社会が求める新たなビジネスの創出できるデータサイエンスのトップレベル人材『データアントレプレナー』を育成 基礎学習: 確率論・統計学、コンピュータサイエンス特論、プログラミング言語特論 対面学習: データアントレプレナー実践論、データサイエンティスト特論 実践学習: インターンシップ、共同研究 補講: データサイエンス基礎講座、EDA(探索的データ分析)とデータサイエンス基礎、Pythonによるデータ集計100本ノック、回帰モデル作成、判別モデル作成、因果推論
【幹事機関】 大阪大学	データ関連人材育成 関西地区コンソーシアム https://duex.jp/	神戸大学・和歌山大学 大阪府立大学・滋賀大学・大阪市立大学・奈良先端科学技術大学院大学・京都大学・大阪公立大学	データ関連技術を高度に駆使しデータサイエンスとの掛け算ができる人材の発掘・育成・活躍促進を一貫して行うことを目的。 A: データサイエンス基礎コース: ビジネス創出や社会問題解決において、課題設定、全体俯瞰力、データ収集・統合力、データ分析、データ解釈の能力を身につける座学講義。 B: データサイエンス実践コース: オープンデータ・企業が公開するデータを活用した問題解決型(実践型) Project Based Learning、インターンシップ、共同研究型研修(実習)プログラム C: 医療データ基礎実践コース: 医師・医療従事者が持つ医療データを分析・解析・解釈し、医療スキル向上や学会発表に繋げる短期集中型研修医療データの収集・統合・分析・解釈の知識・スキルに関するe-Learningと実医療への適用についての座学・ケーススタディなどの研修実施
【協働機関】 東京大学大学院	新領域創成科学研究科 生命データサイエンス人材 育成教育プログラム (DSTEP: Data Scientist Training/Education Program) http://www.cbms.k.u-tokyo.ac.jp/curriculum/dstep.html		ゲノム等の大規模生物データについて、高度な基礎科学力を有し、社会ニーズに直結するデータ解析能力を有する人材の発掘・育成を目的とした博士課程3年間の教育プログラム 1. バイオ機能情報解析学: 国内随一の研究者を講師として招き、最先端のゲノム医学研究について議論を行う。 2. 創薬データサイエンス概論: 大規模計測機器の測定原理を理解する。 3. バイオデータプログラミング演習Ⅰ: 研究に必要な高度なプログラミングを将来的に独学できるような「基礎体力」を養う。 4. バイオデータプログラミング演習Ⅱ: 先駆的に開発され試用が開始されている情報解析ツール群を用い、その利用法について習熟する。 5. 創薬データサイエンス演習: 企業へのインターン、若手研究会への参加、医療機関における医療現場のデータ処理を実地に体験する。 6. ドラッグデザイン特論: バイオインフォマティクスとケモインフォマティクスによる創薬研究の目的・原理・現状を理解することを目標とする。

データ関連人材育成(DuEX) 関西地区コンソーシアム

産官学連携 関西地区コンソーシアム

関西地区8大学

大阪大学 数理・データ科学教育研究センター(MMDS)
データピリティフロンティア機構(IDS)

神戸大学 数理・データサイエンスセンター

和歌山大学 データ・インテリジェンス教育研究部門

京都大学 データ科学イノベーション教育研究センター

滋賀大学 データサイエンス教育研究センター
データサイエンス学部

奈良先端科学技術
大学院大学 データ駆動型サイエンス創造センター

大阪公立大学 数理・データ科学教育研究センター(MMDS)



 **DuEX** Data utilizing Expert program
データ関連人材育成関西地区コンソーシアム

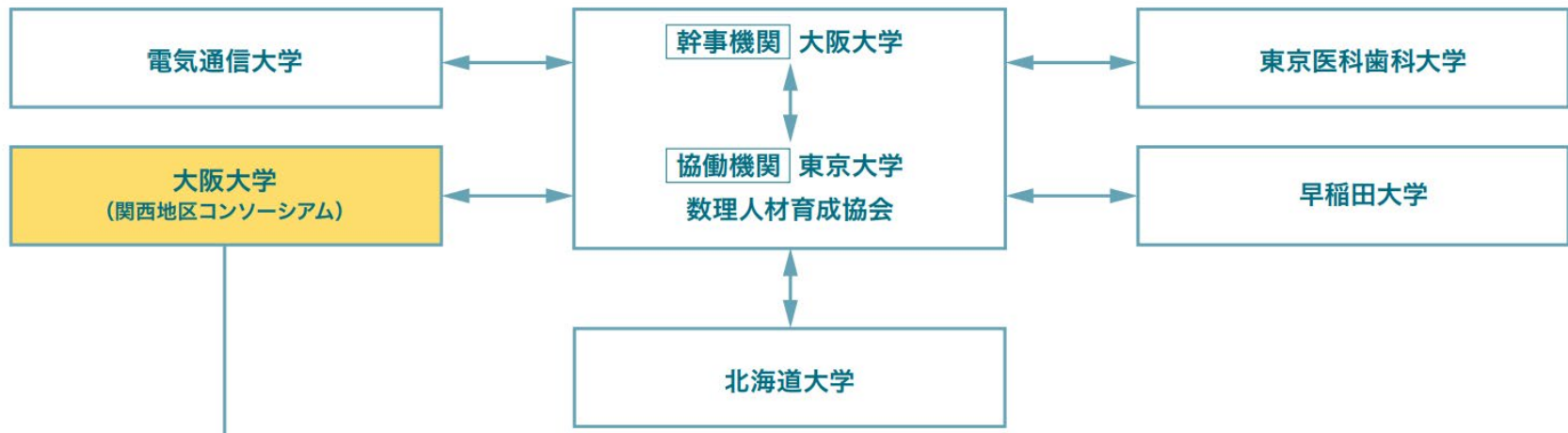
データ関連人材育成(DuEX) 関西地区コンソーシアム 連携機関

- 国立情報学研究所, 統計数理研究所, 産業技術総合研究所, 理化学研究所革新知能統合研究センター, (京都大学)
- 中外製薬(株)、(株)イトーキ、ダイキン工業(株)、(株)オプト、(株)サイバーリンクス、西日本旅客鉄道(株)、(株)リクルートホールディングス、(株)アイディーズ、あいおいニッセイ同和損害保険(株)、第一生命ホールディングス(株)、(株)インテリジェンス、他

平成31年度公募 全国ネットワーク化 【D-DRIVE】

- Doctoral program for Data-Related InnoVation Expert(D-DRIVE)
- 5つのプログラムをつなぎ、全国ネットワーク化

D-DRIVE データ関連人材育成プログラム



大阪大学DuEXのプログラム

- A コース データサイエンス『基礎コース』
 - 対象・修了要件
 - 博士課程学生, 社会人, 数理・情報系の修士課程学生
 - 5単位修得(単位互換協定科目, e-Learning科目)
 - 「基礎コース修了証」
- B コース データサイエンス『実践コース』
 - 対象・修了要件
 - 基礎コース修了(or 同等学力の認定)
 - 実践コース2単位以上修得
 - 「実践コース認定証」
 - 問題解決型(PBL, インターンシップ)授業
- C コース 医療データ『基礎・実践』コース
 - 対象
 - 医学研究科大学院生、研修医、医学系若手研究者
 - 修了要件
 - 「スタディグループ」または「PBL」を含む3単位以上修得。
 - 「医療データコース認定書」





DuEX Data utilizing Expert program

データ関連人材育成関西地区コンソーシアム

構成科目

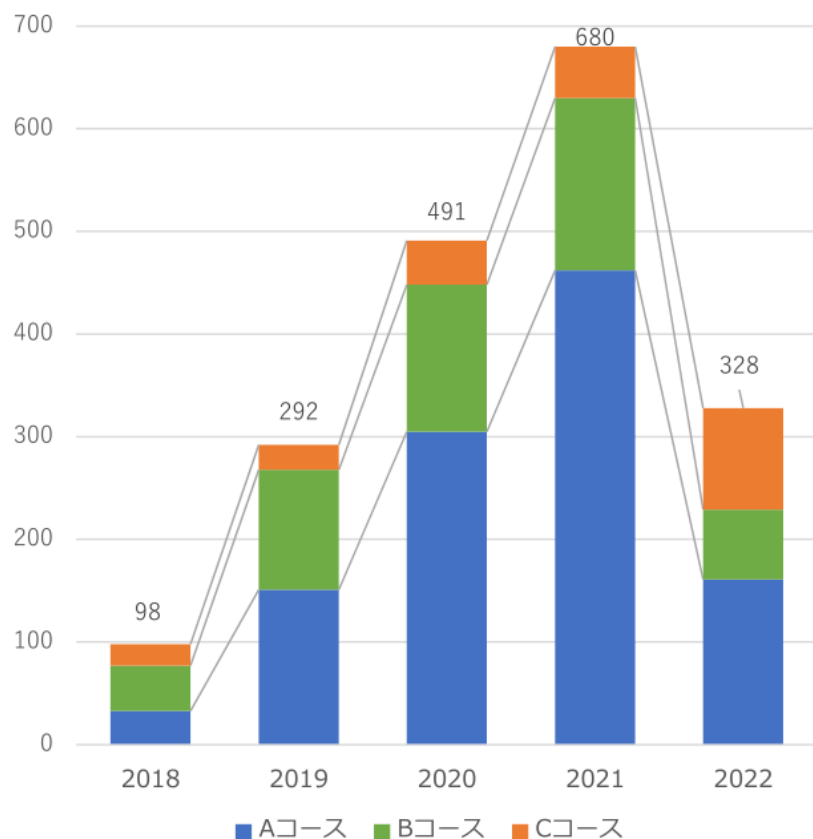
科目分類	科目名	開講大学	単位数	開講区分
数学基礎	データサイエンス概論 終了	滋賀大学	2	講義
統計学	データサイエンス特別レクチャー 終了	滋賀大学	2	講義
数理モデル	モデリング基礎理論 終了	滋賀大学	2	講義
機械学習	データサイエンス特論1 受付中	神戸大学	1	講義
情報学基礎	データサイエンス特論 終了	大阪公立大学	2	講義
機械学習	機械学習特論 終了	大阪公立大学	2	講義
情報学基礎	データマイニング 終了	大阪公立大学	2	講義
プログラミング	機械学習の数理 with R/Python 受付中	大阪大学	2	講義
統計学	多変量解析 受付中	大阪大学	2	講義
統計学	データ科学のための数理 受付中	大阪大学	2	講義
統計学	データ科学と意思決定 受付中	大阪大学	2	講義
機械学習	スパース推定の数理と機械学習への応用 受付中	大阪大学	2	e-Learning
統計学	確率的グラフィカルモデルと因果推論 受付中	大阪大学	2	e-Learning
機械学習	カーネルの機械学習への応用 受付中	大阪大学	2	e-Learning
統計学	情報幾何入門 終了	大阪大学	2	講義
統計学	数理統計入門 終了	大阪大学	2	講義

受講者等

DuEXコース毎受講者数推移

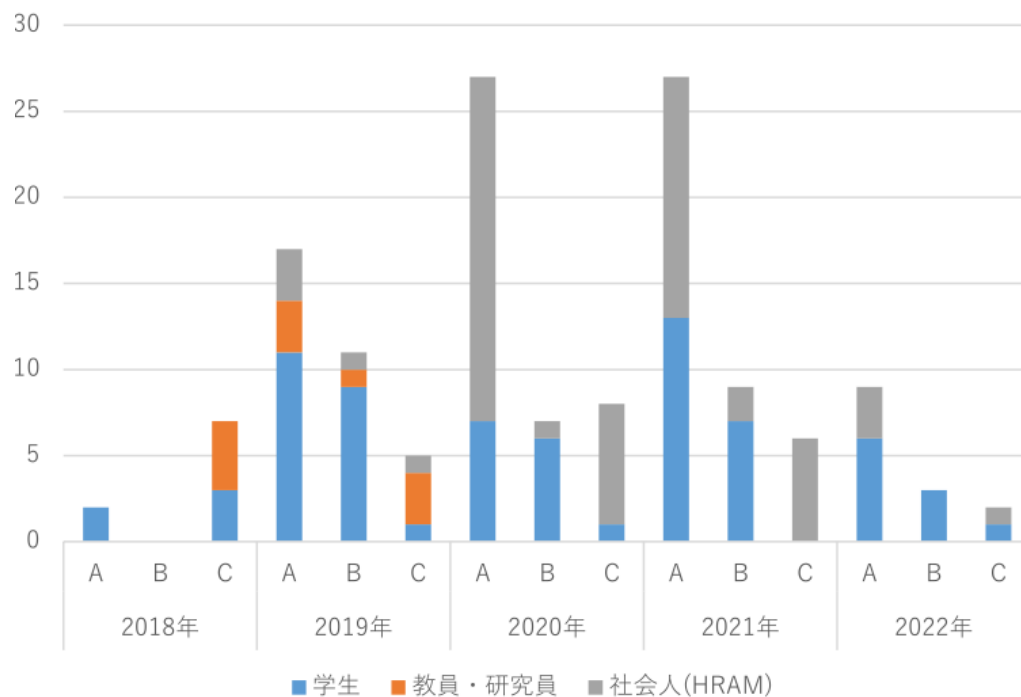
科目履修・e-Learning視聴者
コースの重複を含む

コース受講者数推移(実数)



DuEXコース修了者数推移

2022年度社会人修了者4名
大学院生9名
博士課程在籍者2名



インターンシップ(DuEX)

2022年度インタラクティブマッチング申込企業一覧

企業名
・ 一般社団法人社会デザイン協会
・ ジョンソンエンドジョンソン日本法人グループ
・ フレンズ株式会社
・ 株式会社日立システムズ
・ パナソニックグループ
・ 日本新薬株式会社
・ 株式会社豊田中央研究所
・ 京セラ株式会社

計36社

2022年度インターンシップ実施学生一覧

学生大学	学生専攻	課程	学年	インターンシップ先企業
奈良先端科学技術大学院大学	先端科学技術研究科先端科学技術専攻データサイエンスプログラム	博士前期課程	2年	株式会社豊田中央研究所
大阪大学	生命機能研究科	博士後期課程	3年	岩井コスモ証券株式会社
奈良先端技術大学院大学	先端科学技術研究科	博士前期課程	1年	株式会社フツパー
大阪大学	理学研究科	博士後期課程	3年	小林製薬株式会社
大阪大学	理学研究科	博士前期課程	1年	小林製薬株式会社

【D-DRIVE】全国ネットワーク 全国合同インタラクティブマッチング

第8回

オンライン個別面談・対面交流会会場：大阪大学中之島センター

個別面談2023.6.13(火)～6.23(金)/対面交流会6.24(土)

【企業・研究機関】14社

個別面談参加企業(11社)

- ・株式会社日立システムズ
- ・日東電工株式会社
- ・南海電気鉄道株式会社
- ・関西電力株式会社/K4Digital株式会社
- ・アポロ株式会社
- ・株式会社フツパー
- ・中外製薬株式会社
- ・アクサ生命保険株式会社
- ・株式会社アルトナー
- ・住友電工株式会社
- ・デジタル・アドバタイジング・コンソーシアム株式会社

対面交流会参加企業(12社)

- ・株式会社日立システムズ
- ・京セラ株式会社
- ・日東電工株式会社
- ・サラヤ株式会社
- ・南海電気鉄道株式会社
- ・関西電力株式会社/K4Digital株式会社
- ・アポロ株式会社
- ・株式会社フツパー
- ・住友ファーマ株式会社
- ・株式会社アルトナー
- ・住友電工株式会社
- ・デジタル・アドバタイジング・コンソーシアム株式会社

【学生】27名

東京大学 2名
東京医科歯科大学 2名
大阪大学 10名
滋賀大学 7名
奈良先端科学技術大学院大学 5名
神戸大学 1名

6/13～6/23 企業と学生の個別面談
(110件)

6/24 対面交流会プログラム

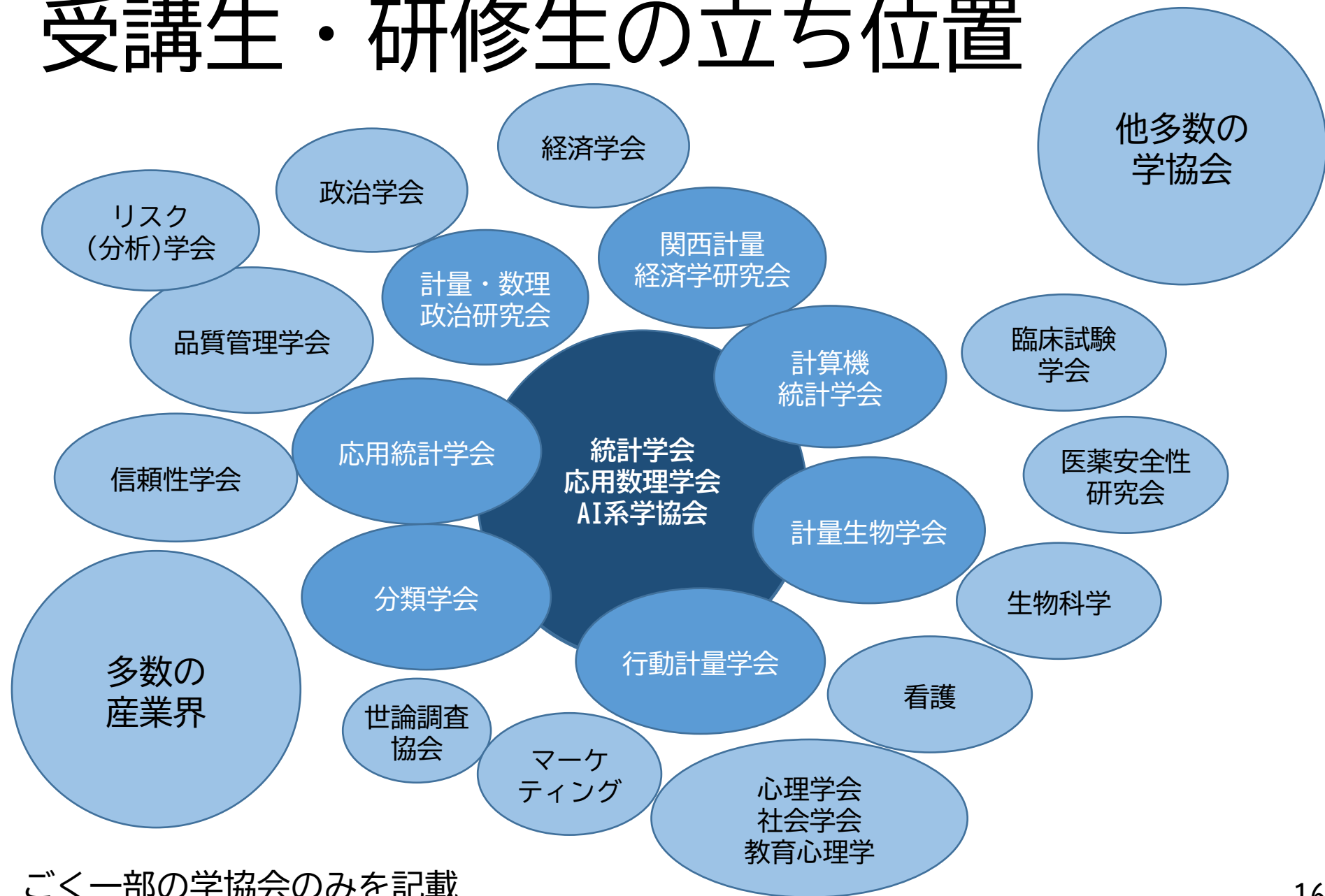
1. 開会式
2. インターンシップ報告会
3. インタラクティブマッチング
企業紹介
4. 閉会式

スタディグループ

- 産業界や諸科学で生じる具体的な課題について、当該分野のエキスパートと共に一定期間集中的に議論する「課題解決型研究集会」
- 【参加団体と課題例】
 - (株)A社
 - 脈波センサを使った体動時脈拍数の検出と…
 - 大阪大学 産学連携本部 イノベーション部
 - 若手研究者のアカデミア就職に関連するパラメータの同定
- 【進行】
 - 初日 午前
 - 課題の説明, ワーキンググループの結成
 - 初日 午後～5日目 午前
 - 集中討議
 - 5日目 午後
 - 各グループ代表者による議論内容と成果の発表
 - 交流会

Point!

受講生・研修生の立ち位置



ごく一部の学協会のみを記載

実施側の視点

高度データ関連人材

- 博士人材
→高度データ関連人材
→多様なキャリア展開
- 修了後
 - アカデミック
 - DS・AIを正しく理解する
 - 民間企業, 実務家
 - 専門研究経験+DS・AI
- 研修内容
 - DSの基礎事項の確認
 - DSの先端知識の獲得
 - インターンシップ・PBL
- 数理・データ・AIの研究者養成は主目的でない

大学統計教員 (統計エキスパート人材)

- PD・若手教員
→高度データ関連人材
→「数・デ・AIを指導できる」後進の養成
- 修了後
 - 主にアカデミック
- 研修内容
 - DSの基礎事項の確認
 - DSの先端知識の獲得
 - 教授するためのスキルの獲得
- 数理・データ・AIの研究者養成は主目的でない
 - 研修生は各専門で活かせるDSを磨く

研修生側の視点

高度データ関連人材

- 自身の専門における数理・DS・AIの理解を深める
- 企業での活躍・実務家への道も
 - 課題解決のための、幅広いデータ分析の知識
 - インターンシップ
 - PBL
 - スタディグループ

大学統計教員 (統計エキスパート人材)

- 自身の研究分野における計量的方法のエキスパート(大学統計教員)に
- 指導者としての準備
 - 系統立てた知識
 - 教えるための知識及びそのテクニック
 - 教材開発

研修生側の視点：教材例

大学統計教員(TeSS)

- 例1 確率変数の導入

- 例2 確率分布の理解

- 正規分布

- 誤差, 身長, 知能テスト
- 様々なファクターの総和

- 指数分布

- 気体分子エネルギー
- 抑うつスコア
- 絶え間ないやり取りの総体
- 偶発故障時間の分布

- べき分布・コーシー分布

- 世帯別金融資産
- 市場の変動
- 外れ値的な値

- 例3 教育教材

- 自身の専門分野におけるデータの活用

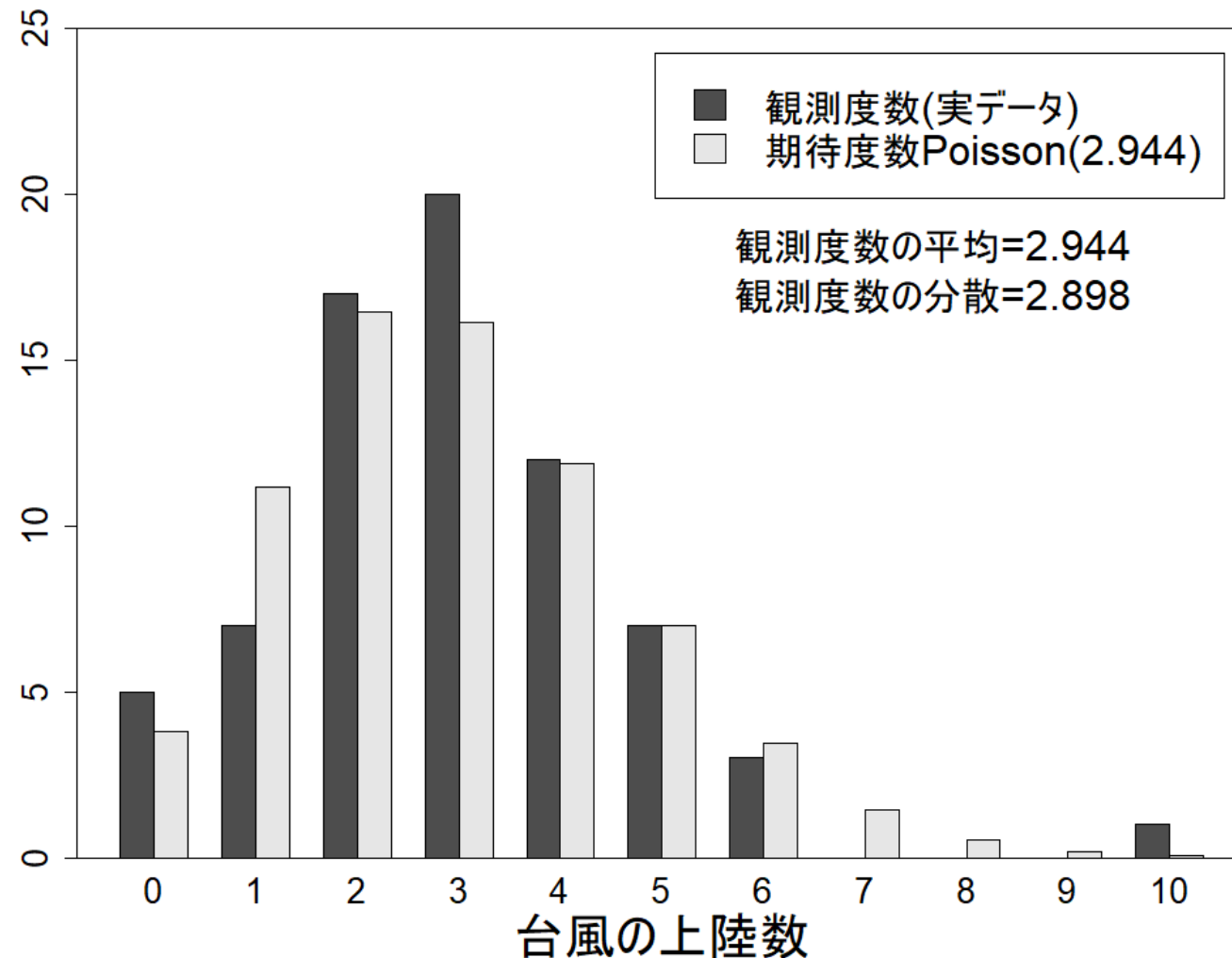
- 単純化
- 目的の明確化

- 身近なデータの利用

- 次スライドへ



1年あたりの台風の上陸数の分布 (1951-2022, n=72)



台風が中心が北海道、本州、四国、九州の海岸線に達した場合を「日本に上陸した台風」としています

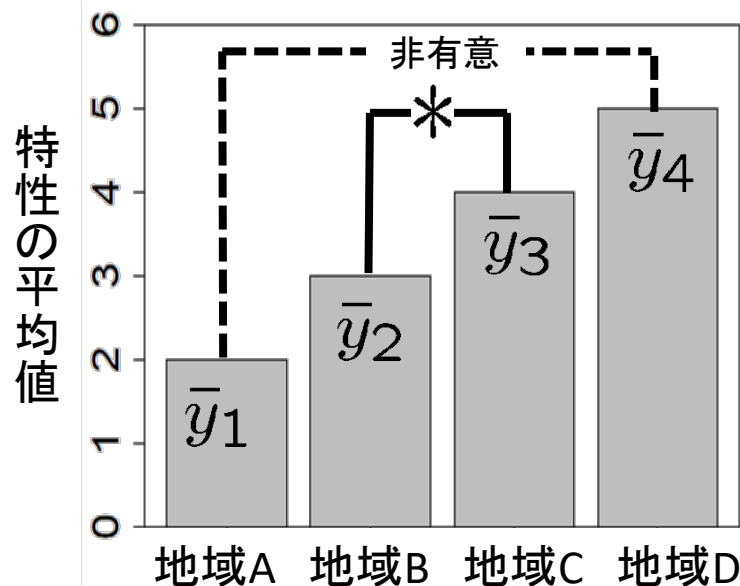
<https://www.data.jma.go.jp/fcd/yoho/typhoon/statistics/landing/landing.html>

<https://newsdig.tbs.co.jp/articles/-/660906>

受講生側の視点：教材例

高度データ関連人材(DuEX)

- 例 反復数不揃いのANOVA
 - (単純)主効果の推測
 - 平均差の大小と検定結果が整合しない
 - 調和平均で n 数を揃える
 - 平均差と検定結果が整合する
- 研修内容
 - 検定・推定のロジック
 - α と β
 - 平方和の分解
 - 確率分布・標本分布
 - 確率的な現象の理解
 - (多重比較)



- Lesson learned
 - ANOVAでは反復数(セルサイズ)を揃えるべく努力する
 - 全てを満足する方法論は無い。メリット・デメリット(リスク)を評価して(苦渋の)選択をする

補遺：t-検定との関係

二標本問題

$$T = \frac{\bar{Y}_1 - \bar{Y}_2}{u\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} = \frac{\bar{Y}_1 - \bar{Y}_2}{u\sqrt{\frac{1}{n_*} + \frac{1}{n_*}}}$$
$$\text{with } n_* = \frac{2}{\frac{1}{n_1} + \frac{1}{n_2}}$$

k 標本問題

$$T_{ij} = \frac{\bar{Y}_i - \bar{Y}_j}{u\sqrt{\frac{1}{n_i} + \frac{1}{n_j}}} = \frac{\bar{Y}_i - \bar{Y}_j}{u\sqrt{\frac{1}{n_*} + \frac{1}{n_*}}} \text{ for } 1 \leq \forall i \neq \forall j \leq k$$
$$\text{with } n_* = \frac{k}{\sum_{\ell=1}^k \frac{1}{n_\ell}}$$

まとめ

- データ関連人材育成プログラム(DuEX)の活動を,
大学教員育成プログラム(TeSS)と比較しつつサマライズ
した
- 類似点
 - 高度データ関連人材の発掘と育成
 - 研修対象は, 諸科学分野においてDSを利用してきた
若手研究者とその卵
- 相違点
 - DuEX: 博士人材のDS・AIへのキャリア開発
 - TeSS: 次世代DS指導者(大学統計教員)の育成
- プログラムの目的に合わせた研修内容
- 今後に向けて
 - DuEXの修了者がTeSSの研修生に
 - 修了生と研修生の相互交流
 - 両プログラム間の情報交換