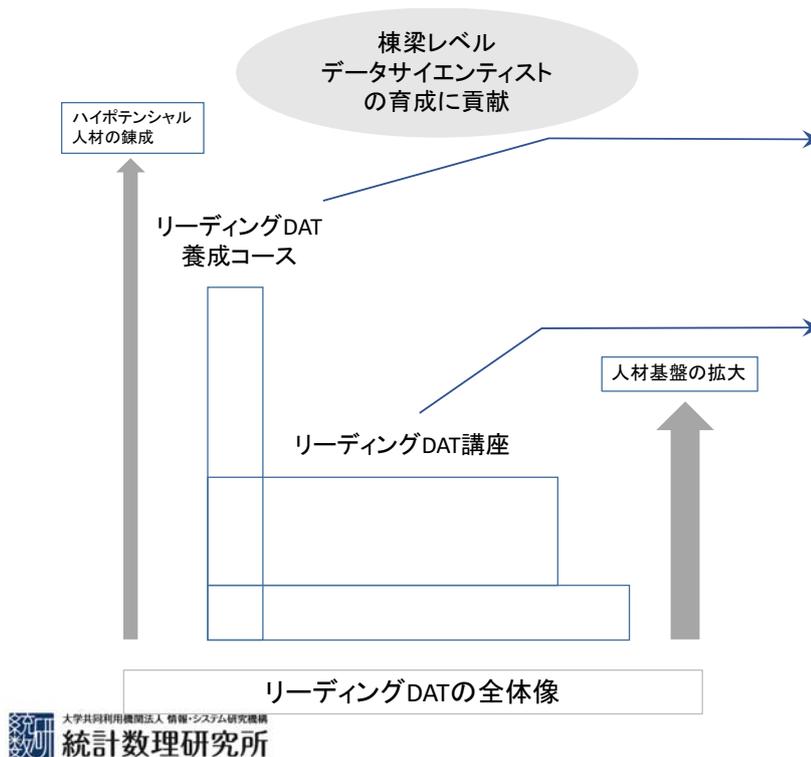


データサイエンス高度人材育成プログラム リーディングDAT (Leading Data Analytics Talents)



ハイポテンシャルな人材を対象として、リーディングDAT講座に加えて実践的な問題の演習や特別講演などを含めた集中的なトレーニング・コース。リーディングDAT養成コースを完了した者には修了認定証を交付。

「棟梁レベル」をめざすデータサイエンティストに必須の統計数理の知識を効率的に習得させることを目的とし、新たに編成された講座。2017年度は「実践ベイズモデリング」(L-B1)と「機械学習とデータサイエンスの現代的手法」(L-B2)の2講座を実施。

0

2017年度リーディングDAT

2018年2月15日・16日開催

L-B1講座 実践ベイズモデリング

「統計モデリングのためのベイズ統計ミニマム」

伊庭幸人 (統計数理研究所)

「GLMから階層ベイズモデリングへ」久保拓弥 (北大・環境科学院)

「状態空間モデル入門 / KFASによる状態空間モデリング」

野村俊一 (統計数理研究所)

「空間データのベイズモデリング」村上大輔 (統計数理研究所)

2018年2月22日・23日開催

L-B2講座 機械学習とデータサイエンスの現代的手法

「イントロダクション」伊庭幸人 (統計数理研究所)

「カーネル法：基礎から最近の発展まで」福水健次 (統計数理研究所)

「因果推論入門」立森久照 (精神保健研究所, 統計数理研究所客員)

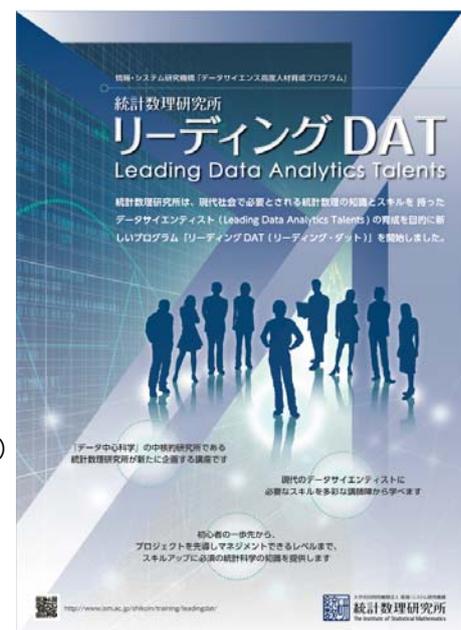
「ニューラルネットワーク入門」庄野 逸 (電通大院・情報理工)

「教師なし学習と行列分解」田中未来 (統計数理研究所)

「スパース推定入門」池田思朗 (統計数理研究所)

2017年度リーディングDAT養成コース

上記2講座受講に加え、2018年3月に課題のレポート講評と特別講演を行い、講座への出席と課題の採点をクリアした受講生には修了認定証を交付。



2017年度実施概要



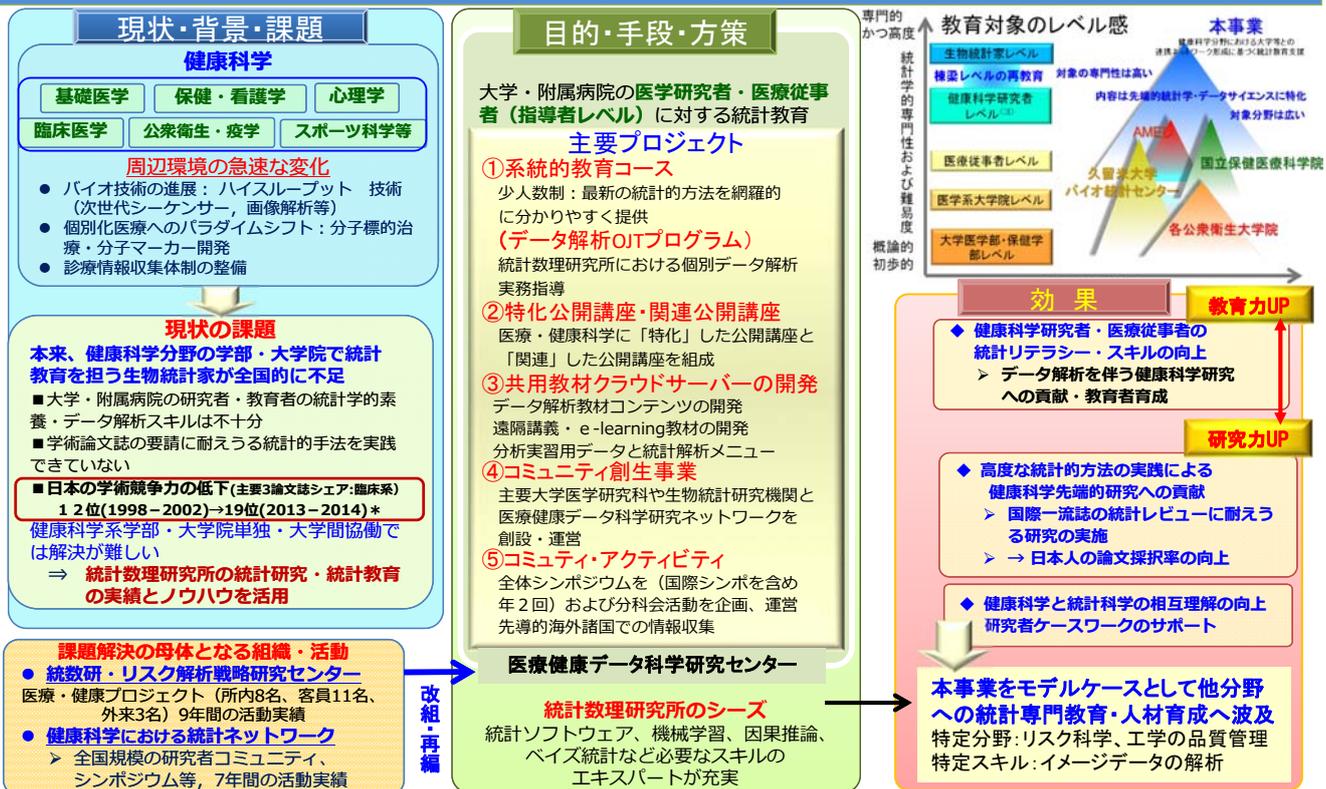
- 「L-B1 実践ベイズモデリング」「L-B2 機械学習とデータサイエンスの現代的手法」の受講登録者はそれぞれ61名。
- リーディングDAT養成コースへの参加者は34名で、規定のコースを修了した25名に修了証を授与。
- 講座、養成コース共に受講者は8割以上が企業からの参加。講義をもとに自習・補習用の教材の整備を行った。

2018年度計画

- ◆ L-A講座「データサイエンスの基礎」を9月19日～21日に開催。
- ◆ L-B1講座、L-B2講座は都心(筑波大学東京キャンパス[茗荷谷])で開催。
- ◆ オンライン教材の拡充



健康科学分野における大学等とのネットワーク形成に基づく統計教育者人材育成および研究・専門性の強化) 概算要求H29から



* 医療産業政策研究所(2015): 政策研ニュースNo.44

系統的教育コース

	生物統計学基礎コース	疫学・公衆衛生統計コース	臨床研究統計コース	生体データ時空間解析コース
イ カ チ イ	野間 久史 (統計数理研究所)	船渡川 伊久子 (統計数理研究所)	立森 久照 (国立精神・神経医療研究センター)	三分一 史和 (統計数理研究所)
内 容	臨床研究・疫学研究で必要となる生物統計学の基礎とEZRを用いたデータ解析について、平易な解説を行う入門レベルのコースです。記述統計、分割表の解析、t検定、Wilcoxon検定から、相関・回帰分析、ロジスティック回帰分析、生存時間解析、サンプルサイズの設定まで、医学研究を行う上で必須となる基礎的な統計手法を習得していただくことが目標です。	疫学・公衆衛生学における統計的側面について学びます。大学や企業で実際に統計解析を行う生物統計学を中心とした医学系研究者を対象とします。疫学・公衆衛生学の概要や研究デザインについて学んだ後、年齢・時代・世代効果、空間統計学、遺伝・分子疫学、因果推論、経時データ解析などの各テーマについて学びます。	臨床研究のデータ解析を行うには、臨床研究で使われる研究デザインの理解が必要です。本コースでは、臨床研究で用いられる研究デザインや基礎概念から始め、臨床研究で使われる実践的なデータ解析の方法までを解説します。その中で「リアルワールドデータ」「傾向スコア」「ビッグデータ」などの現代的トピックも扱う予定です。	生体イメージングデータには観測ノイズ、モーションアーチファクトなど種々のアーチファクトが重畳しており、実際の解析を行う前にアーチファクトを補正するための事前処理が必要不可欠です。本コースでは実際にMATLABプログラミングを行いながら講義をすすめ、データの事前処理法とイベントに関連した有意な生体信号の検出方法の習得を目指します。

医療健康データ科学研究ネットワークと統数研教育事業との関係

統計数理研究所 人材育成 5つの主要活動テーマ

- ・高度専門的教育プログラム(系統別教育コースとOJT)
- ・専門的公開講座
- ・健康科学における統計教育教材開発およびe-learning環境開発
- ・健康科学における統計家の交流シンポジウムおよび分科活動
- ・協力機関の組織化(健康科学研究ネットワーク)

教育ニーズに関する情報
受講者の供給
講師候補者の情報
教材コンテンツの供給

教育事業の提供
教材・Eラーニング環境の提供
研究組織化、活動支援
ネットワーク事務局の運営管理

医療健康データ科学研究ネットワーク

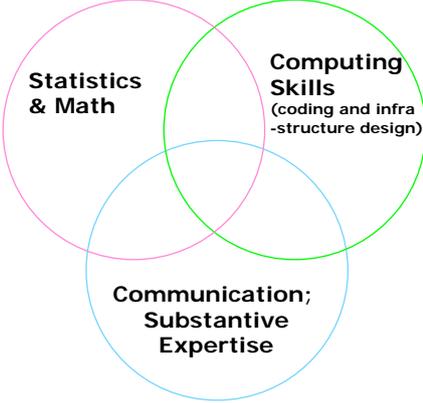
会員組織

運営委員会
会長
運営委員

医療健康データ科学研究
ネットワーク事務局

データ分析コンペティション：実践的な人材育成環境をデザイン

	正解のある問題設定	オープンな問題設定	 <p>Data Analytics Comp</p> <p>What is the expected of data scientists?</p> <p>http://www.ism.ac.jp/shikoin/training/dac/index.html</p>
制限された分析環境	国際大学対抗プログラミング・コンテスト	統数研データ分析ハッカソン 【このハッカソンの他の特徴は、チーム参加、今日的課題、最新分析環境の提供等】	
オープンな分析環境	統数研データ分析チャレンジ 、 The Data Science Bowl, Kaggle, and so on and so forth	データビジネス創造コンテスト、スポーツデータ解析コンペティション等	



Statistics & Math

Computing Skills
(coding and infra-structure design)

Communication; Substantive Expertise

Data Scientists combine several different skills in one profession

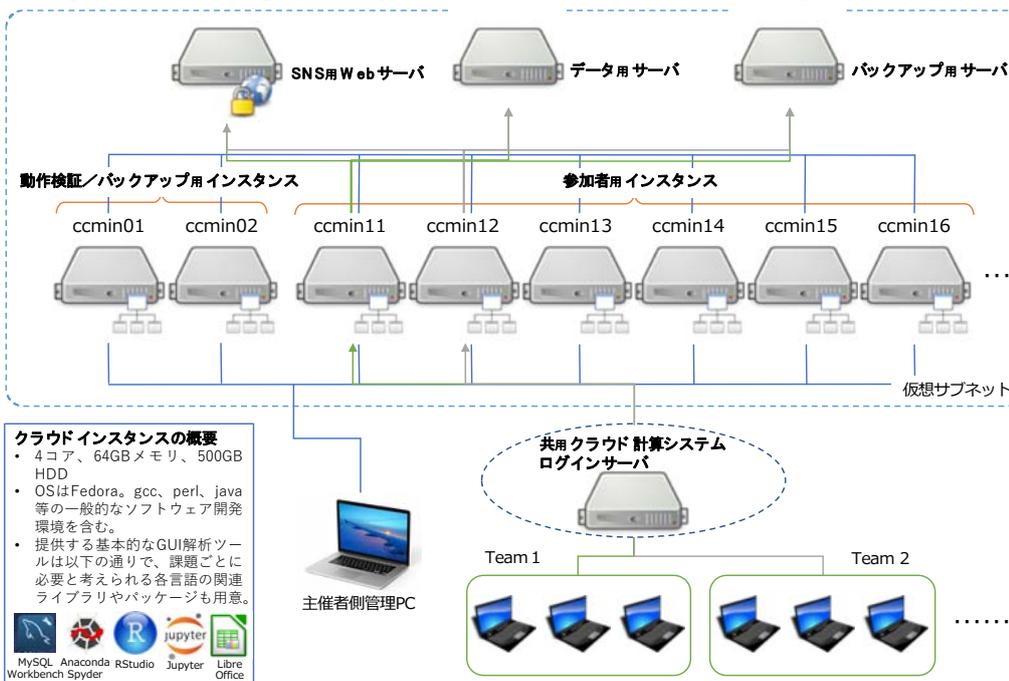
Training of Data Scientists

- For human resources (such as doctoral degree holders) who have already experienced using big data and acquired research methodologies to some extent in various fields of science, **reeducate aimed at career paths for successful participation** in industry and academia, and give a clue for learning necessary.
- Intensive programs for working adults, graduate students and others with practical work experience as data scientists. **Experience cutting edge techniques in the field through PBL, learn the latest accomplishments in various applied domains through case studies.**

参考資料：『ビッグデータの利活用のための専門人材育成について』情報・システム研究機構 ビッグデータの利活用に係る専門人材育成に向けた産学官懇談会, 2015.

← 参考資料：Patil and Mason, Data Driven: Creating a Data Culture, California: O'Reilly Media, 2015など

データ分析ハッカソンのネットワーク構成 ～統計数理研究所共用クラウド計算システムを利用～



クラウドインスタンスの概要

- 4コア、64GBメモリ、500GB HDD
- OSはFedora。gcc、perl、java等の一般的なソフトウェア開発環境を含む。
- 提供する基本的なGUI解析ツールは以下の通りで、課題ごとに必要と考えられる各言語の関連ライブラリやパッケージも用意。




- 参加者の声**
- データ分析は探索的
 - 手法ありき、では上手くない
 - 資源有効利用の重要性
 - リーダーシップの重要性

