

データサイエンス講習会のお知らせ

Rによるデータ解析の基礎と応用

一般社団法人バイオ産業情報化コンソーシアム(JBIC)

東京大学大学院農学生命科学研究科・アグリバイオインフォマティクス教育研究ユニット

2023年1月5日

プログラミング言語・環境であるRには統計やデータ解析に特化した機能がたくさん実装されています。特に学問の世界ではRによるデータ解析が主流ですので、既にPythonなど他のコンピュータ言語を使ったデータ解析を行っている人にも、R言語は身に着ける価値のあるスキルとなっています。本講習会は、これまで(2020年2月、2021年3月、2022年3月)に実施した講習会と同様に、**データサイエンティストの裾野を広げることを目的**として実施します。本年度のデータサイエンス講習会は、**オンライン形式にて2日間開催**します。本年度の講習会は、Rを利用したことない**初心者を対象とした基礎的な内容(1日目)**と**応用的な内容(2日目)**について、Rを便利にするための開発環境であるRStudioを使用した演習形式で行います。

1日目は、R/Rstudioの基本的な使い方・活用方法を学びながら統計解析、データの可視化方法やImageJ Fijiを用いた画像解析に関する基本的な知識と技術を学びます。2日目は、Rによるデータマイニングの基礎的な技術について簡単な演習を行うとともに、近年、生命科学で重要性が高まっている遺伝子型情報を利用したデータ解析手法の一つであるゲノムワイド関連解析(GWAS)について学びます。2日間を通して受講することで**Rを用いたデータ解析を独自で行えるようになることを目指します**。

本講習会では基本的に植物から得られた情報を対象とした解析を行います。受講者各々が取り扱っている生物種にも適用可能な解析方法についての解説・演習となっています。

*本講習会は、2019年2月に、経済産業省の委託事業として実施した『「産業界横断的なバイオ分野データサイエンス関連人材」のスキル標準の策定及びそれらの育成プログラムの開発/実証』バイオインフォマティクス講習会の後継として実施するものです。

URL: <https://www.learning-innovation.go.jp/verify/c0037/>

日時: 2023年3月6日(月)~7日(火)、9:00~16:00(予定) 各日の講義終了後に質問の時間を設けます(参加自由)。

会場: **Webex オンライン講義**

対象者: 講義開催日時点でJBIC会員企業に所属される方

※接続用URLは、受講申込者に、講習会前日までにメールでお知らせいたします。

※講習会終了後、講義動画をJBIC会員サイトにて公開予定です。

1日目(2023年3月6日(月)予定内容):初心者向け講習会(基礎)

・午前:Rを用いた統計解析・データ可視化-入門(東京大学大学院農学生命科学研究科・助教 反田直之)

1日目午前は、Rを用いたデータの読み込みから整形、簡単な統計解析とグラフ化、画像ファイルの保存までの一連の作業を学びます。演習はデータ解析入門者向けの基礎知識の解説を織り交ぜて行います。

・午後:ImageJ Fijiを用いた画像解析-入門およびRでインタラクティブなグラフの作成:(東京大学大学院農学生命科学研究科・助教 反田直之、東京大学大学院農学生命科学研究科・准教授 大森良弘)

1日目午後は、ImageJ Fijiを用いて画像から情報を抽出する技術の基礎を学びます。また、後半にはRでのインタラクティブなグラフの作成について簡単な演習を行います。データの探索や理解を向上させる高度なデータ可視化技術の基礎を学びます。

2日目(2023年3月7日(火)予定内容):初心者～中級者向け講習会(応用)

・午前:Rによるデータマイニング-入門(東京大学大学院農学生命科学研究科・准教授 大森良弘)

2日目午前は、大量のデータから知識を見出すために必要なデータマイニング技術についてRを用いた簡単な演習を行います。Rでデータ解析を行うための基礎的な方法について学びます。

・午後:Rを用いたゲノムワイド関連解析(GWAS)(東京大学大学院農学生命科学研究科・助教 石森元幸)

2日目午後は、近年ヒトや様々な動植物集団における原因遺伝子探索手法として一般的となったゲノムワイド関連解析(GWAS)を取り上げ、Rパッケージ RAINBOWR を使用して、遺伝子型情報(ゲノムワイドマーカー)を活用した解析の演習を行います。GWASは中級者としてはやや難しい内容ですが、生命科学で扱うビッグデータの一つであるゲノム情報の解析について学びます。

受講申込:<https://forms.office.com/r/gCA4J8r4SJ>にある参加申し込みフォームよりお申込みください。

(*本講習会はJBIC会員企業に所属されている方に限定させていただきます。)

申込期間:2023年1月5日(水)～3月3日(金)12:00まで

受講費用:無料

注意事項:

1. 本講習会はデータサイエンス初心者～中級者を対象としておりますが、事前準備として、講習会で使用するPCへのRとRstudioのインストールを受講者各人をお願いいたします。事前準備は4.「令和4年度の講義資料」内の「RとRstudioのインストール」を参照して行ってください。講習会はRとRstudioが問題なくインストールされている前提で行います。
2. 今回のオンライン講義は、RStudioを操作しながら行うハンズオン形式ですので、可能であれば、講義視聴用のPCとRStudio用のPCをそれぞれ別にご用意頂き、講義はデスクトップPCなどの大きめのモニターで視聴いただくことを推奨いたします。
3. 本講習会で利用する遺伝子型情報はバリエーションコール後のファイル(Variant call format;VCFなど)です。FASTQファイルの前処理やマッピングなどは行いませんのでご注意ください。
4. **令和4年度講習会の講義資料**は、以下のJBIC会員専用サイト内に**2月下旬頃**に公開予定です。
https://www.jbic.or.jp/members_only/data_science(ID・PWが不明な場合はJBICにご連絡ください)
5. 両日ともにレポート課題を出します。レポートの提出は必須ではありませんが、提出して**合格と認められれば、「令和4年度R講習会 修了認定証(仮称)」**を授与することを予定しています。

問合せ:

一般社団法人バイオ産業情報化コンソーシアム(JBIC) lecture@jbic.or.jp (奥)
東京大学大学院農学生命科学研究科・アグリバイオ事務局 koshukai@iu.a.u-tokyo.ac.jp (大森)
以上