

データのハンドリング ハンズオンセミナー
～RStudioを使ったデータ操作テクニックを少人数制で徹底指導～

本講座の目標と内容

各回とも、半分以上の時間は R による実習時間といたします。何となく参加したということがないように、講師が受講者と密な双方向のコミュニケーションを取りながらセミナーを進めます。分からない方を置いてきぼりにすることはしません。そのため、10名までの少人数制といたします。

R の基本プログラミングから始め、dplyr や ggplot2 といった R のモダンなパッケージまで学習をおこないます。また、実務でよく用いる、日本語表示に向けたフォントの設定や、Excel ファイルの読み込みなども扱います。ただ、本講座のみでは R によるデータハンドリングを自由自在に行えるレベルに達するのは難しいです。

本講座の最大の目的は、受講者が R の基本を身につけ、自力で R を学んで習熟度を上げていくことができるようにすることです。

<第 1 回>

【目標】 RStudio の設定や R プログラミングの基本を身につける。

◆なぜ R を使うのか

◆RStudio の設定と基本操作

- ・日本語を扱うのに向けたフォント (Myrica) の導入
- ・RStudio の基本操作 (フォントや色の設定なども)

◆R の基本プログラミング

- ・四則演算
- ・データ型 (実数, 整数, 文字列, 論理値, ファクター型など)
- ・ベクトルへのアクセス
- ・論理演算子
- ・データフレームへのアクセス (\$によるアクセス、[]によるアクセス)
- ・データフレームの操作
- ・csv ファイルの読み込み・書出し
- ・繰り返し文 (for 文)
- ・条件分岐 (if 文)
- ・その他 (欠測の処理など)

<第 2 回>

【目標】オープンデータを用いて、`dplyr` と `ggplot2` によるデータハンドリングの基本を身につける（1回目）

- `tidyverse`（Rの強力なデータハンドリングのパッケージ群）の紹介
 - `data.frame` 形式よりも便利な `tibble` 形式
 - パイプ演算子（`%>%`）の紹介
 - `tidyverse` の中でも使用頻度の高い `dplyr` と `ggplot2` の紹介
- `dplyr`（Rの強力なデータ前処理パッケージ）
 - `filter`（値から観測値を選び出す）
 - `arrange`（行を並び替える）
 - `select`（名前を変数を選ぶ）
 - `mutate`（新たな変数を作る）
- `ggplot2`（Rの強力なデータ可視化パッケージ）
 - `geom` による棒グラフや散布図の作成
 - `png` 形式によるグラフ保存

<第3回>

【目標】オープンデータを用いて、`dplyr` と `ggplot2` によるデータハンドリングの基本を身につける（2回目）。また、実務でよく用いる `Excel` ファイルの読み込み、フォルダ内にある同一形式のファイルをまとめて処理する方法、ビッグデータ読み込みに向けた `fread` 関数についても説明する。

- `dplyr`
 - `group_by`（グルーピングをおこなう）
 - `summarize`（多数の値から単一の要約量を作る）
 - `left_join` を中心としたデータの結合
- `ggplot2`
 - `theme` によるグラフ書式の設定
 - 日本語フォントでの表示のさせ方
- ファイル操作関連
 - `readxl` パッケージによる `Excel` ファイルの読み込み
 - フォルダ内にある大量のファイルを処理する方法
 - `fread` 関数（`data.table` パッケージ）の紹介

<R コマンダーのセミナーとの違いについて>

JISA で年 2 回開催している R コマンダーのセミナーとは内容が異なります。R コマンダーも本セミナーで使用する `dplyr` や `ggplot2` と同じく、R の機能を拡張するパッケージソフトの 1 つですので、RStudio と連携して使用可能です。本セミナーを受講後に R コマンダーのセミナーにご参加いただくと統計的な分析手法も合わせて学ぶことができます。逆に、R コマンダーのセミナー受講経験のある方が本セミナーにご参加いただければ、実務で統計的な分析するためのデータを扱う実践的なスキルを身につけることができます。