経営システム分野 3年後期 (7712)
データサイエンス(前半)、担当: 林
~~木曜日 3-4限(10:30-12:00)@21号館 2139講義室~~

多変量解析
～ RとRStudioを用いて ～

林 篤裕 (おもひ領域&社会工学専攻)
@11号館 5階 516室
hayashi.atsuhiro@nitech.ac.jp
052-735-5119 (内線 5119)

データサイエンス(前半、10/1-11/19)

1. データサイエンス(Data Science, DS)って?
   RやRStudio とは、インストール
2. 簡単な統計、データの読み込み
3. 集計、頻度、ヒストグラム、
   散布図、散布図行列
4. 検定
5. 重回帰分析
6. 主成分分析、因子分析
7. レポート課題作成・提出

第1回(10/01): 目次

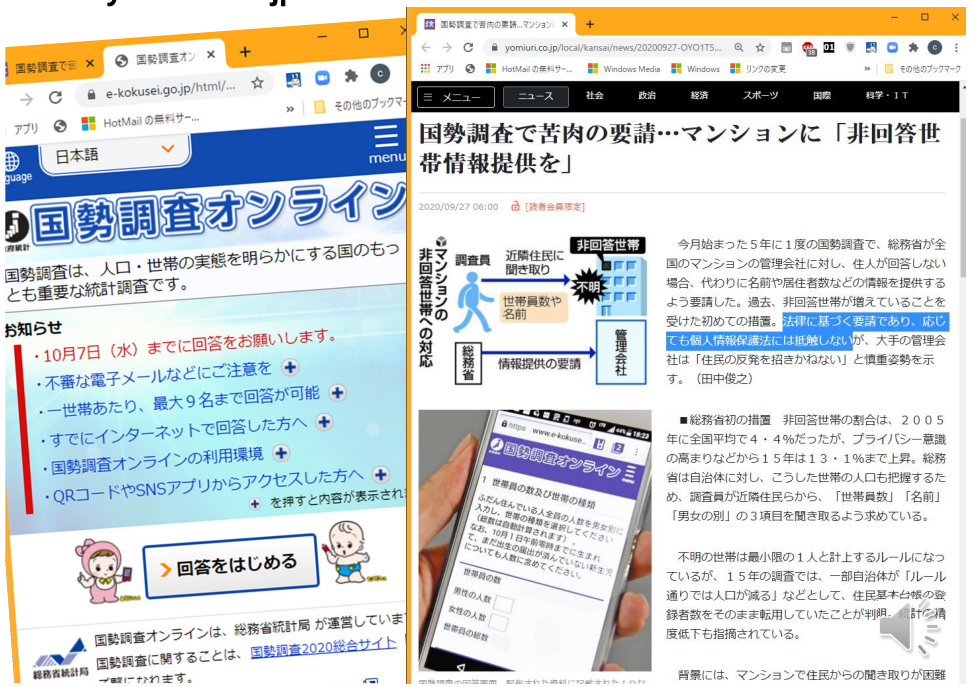
- データサイエンス(Data Science, DS)って?
- 統計って?
- 統計解析に必要なもの
- RとRStudio
  - 各自のPCにインストールしましょう
  - 教育用端末にはインストール済み
- レポート課題(案)の提示
  - 【宿題】データの収集
- アンケート

【提示資料等】Moodleに置いていきます。

1. データサイエンスって?

- Data Science, DS
  - 皆さんのイメージは?
  - 「データサイエンティスト」を目指しておられる方もいらっしゃるかもしれませんね
  - 例えば「国勢調査」は何故行われる? 何で必要? 何のために? 義務だとご存知でしたか?
  - 我々は何でデータを集めるのでしょうか?
- 「データ」を扱う学問としての統計学
  - 情報処理、データマイニング、データサイエンス
  - いろいろな手法、モデルがある
  - データ群の根底に内在する構造(モデル)を明らかに

www.yomiuri.co.jp/local/kansai/news/20200927-OYO1T50006/



1. データサイエンスって?

- Data Science, DS
  - 皆さんのイメージは?
  - 「データサイエンティスト」を目指しておられる方もいらっしゃるかもしれませんね
  - 例えば「国勢調査」は何故行われる? 何で必要? 何のために? 義務だとご存知でしたか?
  - 我々は何でデータを集めるのでしょうか?
- 「データ」を扱う学問としての統計学
  - 情報処理、データマイニング、データサイエンス
  - いろいろな手法、モデルがある
  - データ群の根底に内在する構造(モデル)を明らかに

2. 統計解析に必要なもの

- データを“良く・深く知ろう”とする気持ち
  - データに対する思い入れ
  - 日頃の観察力から育成できるものか?
- 統計手法の修得: 基礎統計量、多変量解析、実験計画法等
- 統計ソフトウェア: 道具
  - Excelではダメなの?
    - 大量データになったら?
    - 複雑な統計手法になったら? 多変量解析...
    - 欠損値の取り扱い
  - 統計向けソフトウエアの利用が一般的 : データ解析
    - BMDP : BioMedical Data Programs(?)
    - SPSS : Statistical Package for Social Science
    - SAS : Statistical Analysis System
    - S, S-PLUS : Statistical
    - R : Sから派生したフリーソフト＋便利ツール
    - LISP-STAT : Lisp で実現、フリーソフト
    - Statistica
    - SAS JMP Pro 14.3
    - Python
    - ...

3. Rの魅力

- フリー(無料)である
- たくさんの開発者が支えている
- 便利ツールも多い
- 本講義では
  - RとRStudioを使う: RStudioはRを支援する環境
  - 本来は教育用端末にインストール済み
  - この状況下では登校を依頼するわけにも行かず
    - 各自のPCにインストールして利用してもらおう
    - インターネット環境が整備できない学生に限り、引き続き教育用端末を設置している講義室(1129, 2139, 2439)の利用が可能です(感染しないように注意を払って)。
- 本日はインストールを完了するところまで

4. RとRStudioのインストール

- ここではWindowsを例に説明するが、MacやLinuxでも同じ環境を手に入れることが可能。
- 以下のWebに丁寧に手順が説明されているので、これに従ってインストールください
  - 「R初心者の館（RとRStudioのインストール、初期設定、基本的な記法など）」
  - https://das-kino.hatenablog.com/entry/2019/11/07/125044
- 現時点での最新版は以下の通り
  - R-4.0.2 for Windows (32/64 bit)
  - RStudio-1.3.1093.exe
- 【作業】上記Webサイトの目次にある以下の2項目を各自で行ってください。次のスライドでは3番目の「・RStudioの機能」から開始します。
  - RとRStudioのインストール
  - RStudioの初期設定

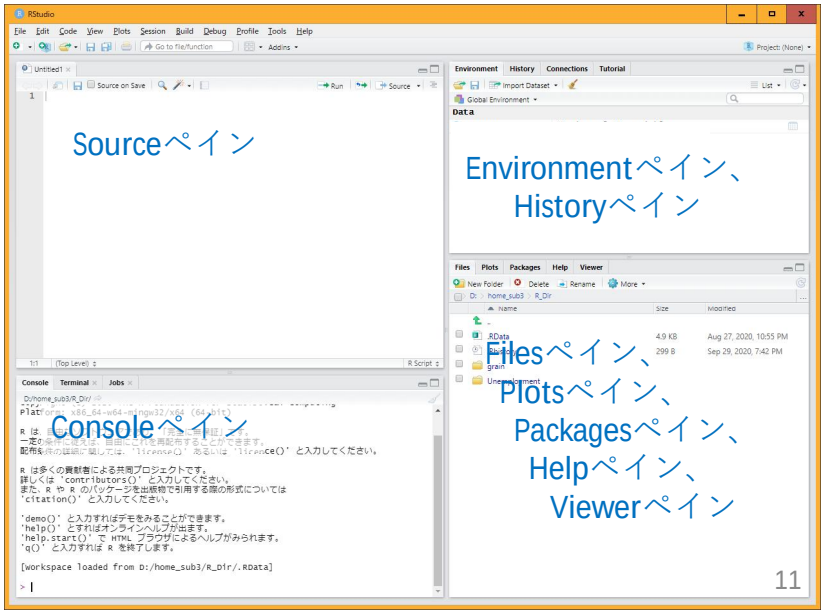
5. インストールできたかの確認

- RStudioの起動
  - デスクトップに表示されているrstudioのアイコンをクリックしてRStudioを起動する
  - Rはrstudio内から操作されるので意識する必要なし

rstudio.exe

- 画面の説明: 4つの画面がありそれぞれに役割あり
  - 左上: Sourceペイン　<=== プログラムを入力する
  - 左下: Consoleペイン <=== 計算結果が表示される
  - 右上: Environmentペイン、Historyペイン
  - 右下: Filesペイン、Plotsペイン、Packagesペイン、Helpペイン、Viewerペイン
- もし、画面が3つしか表示されておらずSourceペインが1行しか表示されていなかったら、Consoleペインの上部(バー部分)をダブルクリックする。

5. インストールできたかの確認



5. インストールできたかの確認

- 手始めに簡単な操作(電卓のような使い方)
- Sourceペイン(左上の領域)で
  - 「3+5」と入力後、「Ctrl+Enter」(Ctrlキーを押しながら、Enterキーを押す)を行うと、Consoleペインに「8」と計算結果が出力される。
  - 「5^3」と入力後、「Ctrl+Enter」を行うと、Consoleペインに「125」と計算結果が出力される。
- 演算記号: 四則演算: ＋、－、＊、/
  べき乗: ^　剰余: %%　平方根: sqrt()
- 「式を入力」後、「Ctrl+Enter」で実行する。
- これでは、まるで単なる電卓
- 来週からより高度な計算について紹介します。

6. 8回を通しての宿題

【レポート(案)】「ご自身で興味を持ったデータ」を分析して報告してもらう。

- 8回の最後には、各自で持ち込んだデータの分析を行ってもらいレポートとして提出してもらおうと考えている。そこで「ご自身の興味のあるデータ」を見つけてきてほしい。電子化し、分析・報告してもらおうと考えているが、持参後の手順等は今後順に紹介するので、まずは「興味あるデータ」を各自自力で見つけてきてほしい。習慣にすると良いかも。
- 6-7回が終わる頃までには確定すること。そうしないとレポートが書けない＝単位を出せない。

7. アンケート(ショート課題)

- 以下項目について「10月2日夕方まで」に回答下さい。本日(第1回)の出欠調査を兼ねます。
- 送信先: stat.nitech@gmail.com
- メールの表題(件名)は以下のフォーマットで
  - [DS20] Yamada

a. 所属、学年、学籍番号、氏名
b.「データサイエンス」に抱くイメージや印象
c. 講義についての要望等
d. [データ収集] 身長、体重、胸囲、自宅生/下宿生の別、仕送り額、スマホの月額通信料(概数)
e. その他、質問やお気づきの点があれば何なりと。