

教育現場におけるJDCat分析ツールの活用

学術情報基盤オープンフォーラム2024「SINETStream・研究データ解析基盤」

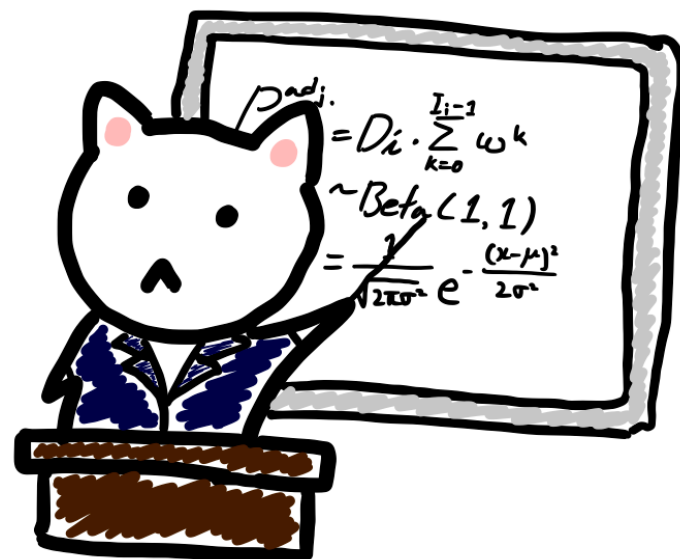
そん じえひょん

宋 財 沅

関西大学総合情報学部 准教授

2024-06-12

自己紹介



プロフィール

- そん じえひょん 宋 財 沄 (SONG JAEHYUN)
 - 関西大学総合情報学部 准教授
 - 博士（政治学）
- 専門は政治行動論、選挙研究、政治学方法論
 - 情報学、計算機科学、データサイエンスではありません。
- 趣味はゲームとラーメン屋巡り
 - 急募）神保町・東京駅周辺のラーメン屋

私とNIIとJDCat分析ツール

- NIIとの接点はございません…。
- JDCat分析ツールを授業に導入してから4年目（2021年度～）
- 関西大学、神戸大学、京都府立大学、高知工科大学（予定）

JDCat分析ツールの活用事例（関西大学）

関西大学とJDCat分析ツール

- 「ミクロ政治データ分析実習」 （前期）
 - **Rの使い方が中心**
 - データ入出力、前処理、可視化など
 - 履修者：約40人（3回生以上） + TA2名
- 「マクロ政治データ分析実習」 （後期）
 - Rを使った**データ分析**が中心
 - リサーチデザイン、統計的仮説検定、検定、相関分析、検定、回帰分析など
 - 前期科目の履修を強く**推奨**しており、強制はできない。
 - 履修者：約30人（3回生以上） + TA2名
- 文理融合学部であるため、Python（半数程度）、JavaScript（全員）の経験者も多い

使用ツール・リソース

1. JDCat分析ツール

- IDEはJupyter Notebookを使用せず、R定番のIDE「RStudio Server」を使用
- 分析環境のレポジトリ：https://github.com/JaehyunSong/Binder_R

2. 講義のサポートページ：<https://www.jaysong.net/r4ps/>

- シラバス、R環境のセットアップ、講義スライド、講義ノート、課題案内

3. 関大LMS

- 主に実習用データの配布、課題提出窓口として使用

4. 参考書（教科書なし）

- 前期：『私たちのR：ベストプラクティスの探求』（<https://www.jaysong.net/RBook/>）
- 後期：『Rによる計量政治学』

課題の様子

Quartoを使用し、教員が作成したサンプルファイルの内容を再現する内容

- 試行錯誤しながら、教員と同じ結果を出す。

****問題11:**** `df1`の`Education\_C`の値ごとに`Ideology`の平均値、標準偏差、最小値、最大値を計算し、それぞれ`Mean`、`SD`、`Min`、`Max`という列として出力する。各グループに属するケース数も`N`という名の列として出力すること。出力される結果に置いて`Education\_C`列の名前は`Education`に修正する。以上の結果は`df2`という名のオブジェクトとして格納し、`df2`を出力すること。

```
```{r}
ここにRコード
```
```

****問題12:**** `df1`の性別（`Gender\_C`）ごとの世帯収入（`Income`）の平均値、標準偏差、最小値、最大値を計算し、それぞれ`Mean`、`SD`、`Min`、`Max`という列として出力する。各グループに属するケース数も`N`という名の列として出力すること。出力される結果に置いて`Gender\_C`列の名前は`Gender`に修正する。以上の結果は`df3`という名のオブジェクトとして格納し、`df3`を出力すること。

```
```{r}
ここにRコード
```
```

Problem set

問題11: `df1`の`Education_C`の値ごとに`Ideology`の平均値、標準偏差、最小値、最大値を計算し、それぞれ`Mean`、`SD`、`Min`、`Max`という列として出力する。各グループに属するケース数も`N`という名の列として出力すること。出力される結果に置いて`Education_C`列の名前は`Education`に修正する。以上の結果は`df2`という名のオブジェクトとして格納し、`df2`を出力すること。

ここにRコード

```
# A tibble: 3 × 6
  Education Mean    SD  Min  Max    N
  <fct>      <dbl> <dbl> <dbl> <dbl> <int>
1 Low       5.58  2.11    0    10   156
2 Middle    5.41  1.93    0    10   654
3 High      4.98  2.06    0    10   201
```

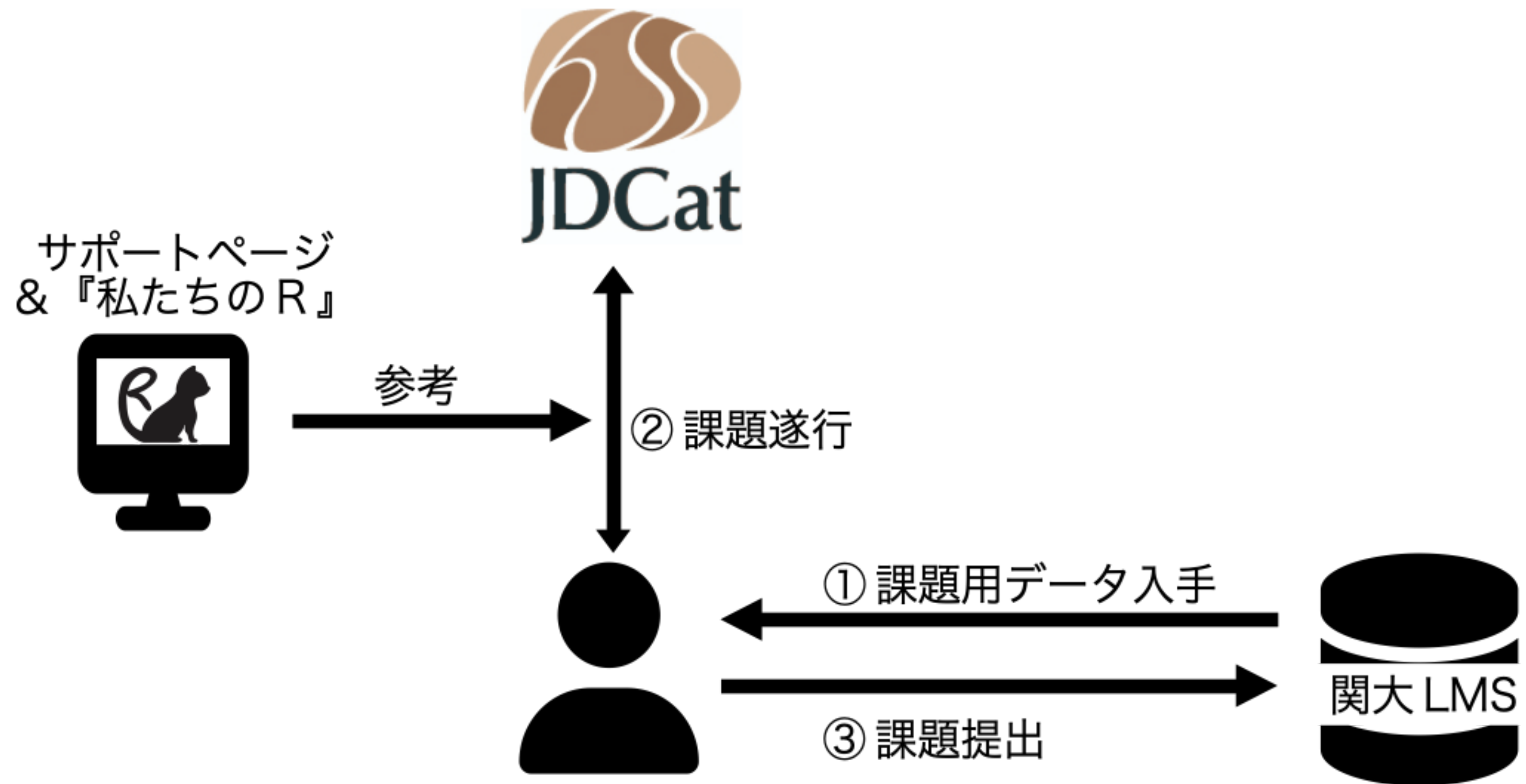
問題12: `df1`の性別（`Gender_C`）ごとの世帯収入（`Income`）の平均値、標準偏差、最小値、最大値を計算し、それぞれ`Mean`、`SD`、`Min`、`Max`という列として出力する。各グループに属するケース数も`N`という名の列として出力すること。出力される結果に置いて`Gender_C`列の名前は`Gender`に修正する。以上の結果は`df3`という名のオブジェクトとして格納し、`df3`を出力すること。

ここにRコード

```
# A tibble: 2 × 6
  Gender Mean    SD  Min  Max    N
  <chr>   <dbl> <dbl> <dbl> <dbl> <int>
1 Female 4.40  2.55    1    10   499
2 Male   4.36  2.45    1    10   512
```

サンプル

課題遂行の流れ



JDCat分析ツールのメリット

クラウド環境が持つ強み

1. 履修者全員が同じ環境で実習が可能

- 教員が予め動作を確認しておけば、動作は100%保証される。
- もちろん、サーバー側の問題、キャンパスのネットワーク環境により、全員ダメになる可能性もある。場合によってはローカル版のR + RStudioへの切り替えも考えておく必要がある。

2. (キャンパスー自宅間の) 学習体験の均質化

- 自宅のPCにRのインストールはハードルが高い、場合によってはスペックが要求される場合も
- YouTubeが見れる程度のPC・~~スマホ・タブレット~~であれば、キャンパスでの学習体験がそのまま再現可能

3. Rパッケージのインストール、アップデートが比較的自由

- 大学のPCだとプロキシサーバーという邪悪な存在により、パッケージのインストールが出来ないケースが多い
- ただし、`apt` が使えないので、ライブラリが求められてしまうとそのパッケージはアウト
 - 宋が作成したレポジトリはよく使われるRパッケージのインストールには困らないように設定はしている。

4. NIIの対応の素晴らしさ

JDCat分析ツールへの要望

JDCat分析ツールへの要望

門外漢なので好き勝手に言わせていただきます。

1. 分析環境の共有

- （理想としては）アカウントをグループ化し、教員アカウントは当該グループ内の（特定の）分析環境にアクセス可能に
- または、コードの共有プラットフォーム
 - 各スレッドを作成者、教員、TAのみ閲覧・コメント可能な掲示板程度でも良い。
 - 質問内容入力欄、コードのコピペ欄（line number & syntax highlighting）
 - AI・DS教育用プラットフォームへの発展
- 授業外における質疑応答の効率が大幅に上昇

2. サーバーリストの改善

- サーバー名の識別力が低く、分析環境を使い分ける人にとっては不便

授業外のフィードバック

- 対面での質問を嫌がる学生が大半で、テキストベースのフィードバックが多い（非効率的）
 - メール、メッセージ、掲示板、Discordチャット
 - 対面（授業中、授業外のZoom/Discord）での質問を推奨しているものの、利用率は極めて低い

< LINE



2023年6月4日



質問があります。情22-1234 の
NOBUNAGA ★O★D★A★です。
問10ですが、以下のエラーが出ます。

23:12



ERROR: Hi, my name is Taro.
Nice to meet you!

23:16

それだけだと微妙ですね、コードも見せてもらえますか？

23:16

```
49 df |>
50   summarise(Rate = mean(rate, na.rm = TRUE),
51             N     = n(),
52             .by   = station) |>
53   arrange(desc(Rate)) |>
54   print(n = Inf)
55
```

23:30

問題なさそうですね。多分、前処理の問題かな？問7のコードも見せてください。

23:31

<https://www.jaysong.net/>

```
Takatsuki <- tibble(name, rate, station, review, bookmark) |>
  separate(col = station,
    into = c("station", "temp"),
    sep = " / ") |>
  select(-temp) |>
  separate(col = station,
    into = c("station", "distance"),
    sep = " ") |>
  mutate(distance = str_remove(distance, "m")) |>
  mutate(across(c(rate, distance, review, bookmark), ~as.numeric(.x)))
```

23:45

ここも大丈夫かな。なら問5は?

23:46

```
name <- raw |>
  html_elements(".list-rst__rst-name-target") |>
  html_text2()
```

23:55

うーん。分かりませんね。ソースコード丸ごと送ってください。

23:56

今日



分かりました。

00:01



私のコードを喰らえ.R

00:20

ありがとうございます。確認してみます。

00:21

<https://www.jaysong.net/>

サーバーリスト

- 複数の分析環境を立ち上げた場合、管理が難しい。
 - 分析環境の生成日時でも確認できたら少しは楽になるかも

Start My Server

Named Servers

In addition to your default server, you may have additional 10 server(s) with names. This allows you to have more than one server running at the same time.

| Server name | URL | Last activity | Actions |
|--|---|----------------|----------------------|
| <input type="text" value="Name your server"/> Add New Server | | | |
| jaehyunsong-binder_r-di28doov | /user/t210047@kansai-u.ac.jp/jaehyunsong-binder_r-di28doov | 21 days ago | stop |
| jaehyunsong-binder_r-dsafvj21 | /user/t210047@kansai-u.ac.jp/jaehyunsong-binder_r-dsafvj21 | 21 hours ago | stop |
| jaehyunsong-binder_r-asdf1234 | /user/t210047@kansai-u.ac.jp/jaehyunsong-binder_r-asdf1234 | 3 hours ago | stop |
| jaehyunsong-binder_r-zxcv4321 | /user/t210047@kansai-u.ac.jp/jaehyunsong-binder_r-zxcv4321 | 3 days ago | stop |
| jaehyunsong-binder_r-bbas28tu | /user/t210047@kansai-u.ac.jp/jaehyunsong-binder_r-bbas28tu | 15 minutes ago | stop |