

データ分析演習 第8回 Python: ここまでの復習 (教科書 ch.2)

兵庫県立大学 社会情報科学部

山本 岳洋

t.yamamoto@sis.u-hyogo.ac.jp

本日の課題: <https://classroom.github.com/a/PaJNDPCe>

レポート3: <https://classroom.github.com/a/AD6LD1A3>

課題の提出状況について

2

- ユニバーサルパスポートに掲示しましたので各自確認しておいてください
 - 毎週木曜日の講義前くらいに更新する予定です

今週とレポート3の目的

3

- 先週・先々週学んだ pandas と matplotlib を用いて実際のデータを分析してみる
 - 簡単な分析であれば，とても短いコードで分析から可視化までできてしまうことが伝わると良い

今週とレポート3で扱うデータ

4

- 総務省が行っている家計調査のデータ
 - 期間: 2010年1月～2019年12月 の10年分
 - 対象: 全国の二人以上の世帯
 - セルの値: 各品目について1世帯あたり1ヶ月間でいくら支出したか（円）
 - 品目: 山本が選定した45品目

week08data.csv の一部

| time | year | month | rice | bread | noodle | fish_faw | fish_salt | fish_process | beef | pork |
|-----------|------|-------|------|-------|--------|----------|-----------|--------------|------|------|
| 2010/1/1 | 2010 | 1 | 1815 | 2255 | 1369 | 3992 | 1140 | 718 | 1579 | 2002 |
| 2010/2/1 | 2010 | 2 | 1980 | 2231 | 1281 | 3720 | 1089 | 660 | 1361 | 1943 |
| 2010/3/1 | 2010 | 3 | 2248 | 2518 | 1447 | 4139 | 1275 | 701 | 1517 | 2009 |
| 2010/4/1 | 2010 | 4 | 2259 | 2468 | 1364 | 3818 | 1181 | 662 | 1462 | 1923 |
| 2010/5/1 | 2010 | 5 | 2282 | 2507 | 1504 | 3973 | 1218 | 660 | 1575 | 1977 |
| 2010/6/1 | 2010 | 6 | 2285 | 2402 | 1612 | 3674 | 1152 | 602 | 1349 | 1872 |
| 2010/7/1 | 2010 | 7 | 2183 | 2365 | 1895 | 3518 | 1203 | 603 | 1504 | 1909 |
| 2010/8/1 | 2010 | 8 | 2114 | 2250 | 1680 | 3772 | 1164 | 607 | 1606 | 1953 |
| 2010/9/1 | 2010 | 9 | 2927 | 2237 | 1280 | 3697 | 1069 | 642 | 1442 | 1973 |
| 2010/10/1 | 2010 | 10 | 3305 | 2381 | 1384 | 3791 | 1122 | 741 | 1534 | 2108 |

出典: 政府統計の総合窓口(e-Stat)「家計調査-家計収支編-二人以上の世帯-月次-品目分類-品目分類（2020年改定）（総数：金額）」
（<https://www.e-stat.go.jp/dbview?sid=0003343671>）より地域区分:全国，2010年1月から2019年12月のデータをe-Statから取得。
データの編集について: e-Statより取得したデータから各品目の支出金額のみを抽出し，品目名を英語名に山本が修正。

対象45品目一覧

(labelinfo.xlsxに記載しています)

5

米 パン 麺類 生鮮魚介 塩干魚介 魚肉練製品 牛肉 豚肉 鶏肉
ハム ソーセージ ベーコン 牛乳 乳製品 卵 キャベツ
ほうれんそう はくさい ねぎ ブロッコリー もやし
じゃがいも だいこん にんじん たまねぎ たけのこ なす
トマト ピーマン りんご みかん オレンジ 梨 ようかん
まんじゅう カステラ ケーキゼリー プリン せんべい
ビスケット スナック菓子 キャンデー チョコレート
アイスクリーム・シャーベット

参考: 家計調査の標本について

6

B - 3 約8,000世帯（二人以上の世帯）の標本は、地域（168層）別にどのように配分されているのですか？

まず、目標とする標準誤差率を基に、全国結果を集計するのに必要な標本サイズである約5,000世帯を各層に配分します。

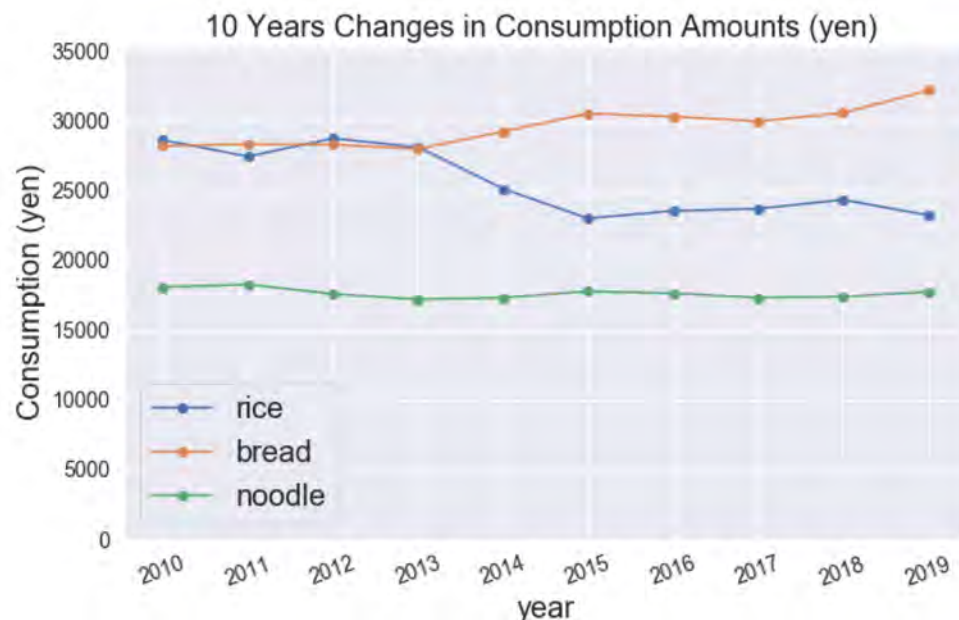
この後、地方別結果を集計するために、標本サイズが約400世帯を下回る地方について約400世帯となるように追加配分します。また、都道府県庁所在市及び政令指定都市別の結果を集計するために、標本サイズが約100世帯を下回る都道府県庁所在市等について約100世帯となるように追加配分します。この結果、全国で約8,000世帯となります。

引用: 総務省統計局 - 家計調査に関するQ&A(回答)

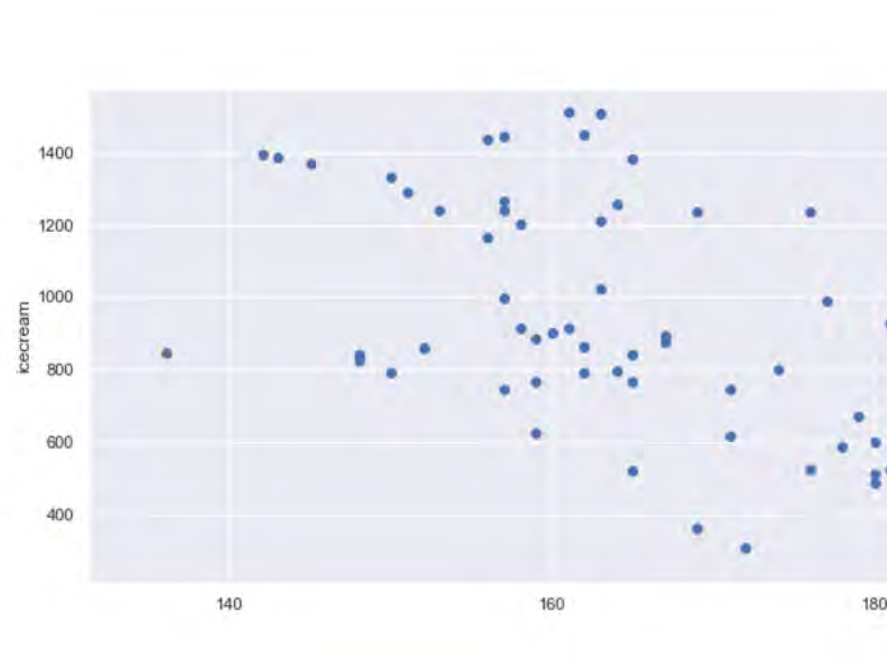
<https://www.stat.go.jp/data/kakei/qa-1.html> (2020年6月22日閲覧)

例えばこんな図を今日の演習で作ります

7



米・パン・麺類の10年間の支出金額の推移



キャンデー(x軸) とアイスクリーム・
シャーベット (y軸) との支出金額の散布図

レポート3

8

- URL: <https://classroom.github.com/a/AD6LD1A3>
- 締切: 7月9日 14:30
- 詳細はノートブックを確認
- 内容は、今週扱うデータの分析です
 - 穴埋めにはしていませんので、今日の資料や教科書を見ながらコードを書いていってください
 - 任意の自由課題もありますので、余裕がある人は積極的に、各自でいろいろな関数を調べながら取り組んでみてください

ありえそうな分析

9

- 日本人の米離れは本当か？
- ここ10年で支出金額が減っている品目は何？
- 消費税増税の影響が表れている品目はあるか？
 - 参考: 2014年4月:5%→8%, 2019年10月: 8% → 10%
- 野菜の旬は支出金額に表れるのか？
- 冷夏な年とそうでない年でアイスの支出金額は異なるか？
 - 今回のデータだけではこれはできないが、気温のデータがあればできそう

今週の演習資料の開き方

10

1. GitHub classroom のリンク
<https://classroom.github.com/a/PaJNDPCe>
2. GitHub Desktop で clone する
3. Show in Explorer でフォルダを開き，
week08_JBxxxxxx.ipynb の名前を変更して自分の学生番号を入れる
4. AnacondaからJupyter Notebookを起動し，2で clone したフォルダから，3で名前を変えた week08_JB…。ipynb を開く

Discordに参加してみる

その前に: 教員への質問について

12

- これまで通り， まずはslackで教員にDMしてください
- Discordが使えて，音声や画面共有した方が早そうであれば slackで質問を受けた後Discordを使うこともあるかも

参加時にマイクがオンになると思ってください

(もちろん後でオフにできます)

Discordとは

14

主にボイスチャットを目的としたコミュニケーションツール。
自由に移動できるZoomのブレイクアウトセッションに近い。

Discordを試す目的: 孤独感の低減

- 学生同士のコミュニケーションとして使ってもらえたら最高に嬉しい
- 明示的にコミュニケーションを取らなくても、お互いがそこにいることの可視化

注意その2

15

- 本日もし講義に関する全体のアナウンスや補足があれば **slackに流します**
- セキュリティの厳しいネットワーク（例:学内）だと音声や映像が届かない，届けられない場合があります
 - 学内の方は **eduroam** から参加してください

手順1: 招待リンクをクリックして名前を入力

16

1. Discordの招待リンク（slackに張ります）をクリック



The screenshot shows a Discord invitation interface. At the top is a circular profile picture of a man with glasses. Below it, the text reads '山本 岳洋[教員]に招待されています' (Invited by Takahiro Yamamoto [Instructor]). The server name is 'SIS-データ分析演習2020'. Below the server name, it says '●1人がオンライン' (1 person online) and '●5人' (5 people). The 'ユーザー名' (Username) field contains 'JB19S123 山田太郎'. A blue button labeled 'はい' (Yes) is below the username field. At the bottom, there is a question '既にアカウントをお持ちですか?' (Do you already have an account?) and a link to the Discord Terms of Service and Privacy Policy.

ユーザ名は

学籍番号 + 名前

手順2: アカウントは作成しなくてOK

17

- アカウント作成画面がその後でますが，無視でOK



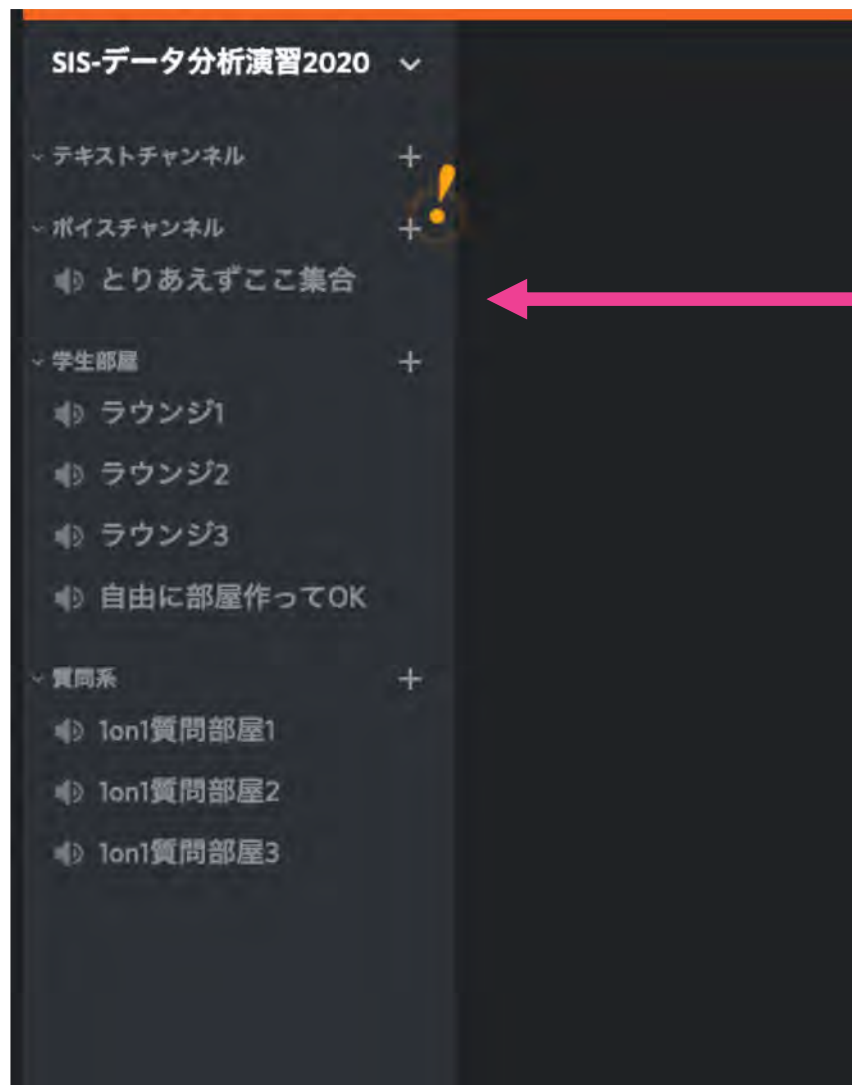
手順3: ブラウザのマイク使用を許可してください (この画面は出ない人も多いかも)

18



手順4: 「とりあえずここ集合」に行く

19



「とりあえずここ集合」という部屋があるので、クリックして参加

山本の音声聞こえたら成功！

ちなみに、とりあえずここ集合部屋は学生はマイクをオンにできないようにしています。

手順5: 後は自由に移動して使ってください

20

- デフォルトで学生部屋を4つ作っています
- 部屋も皆さんで自由に作れます
- マイクオン，画面共有，カメラオンなど自由にできます
 - 意図せずマイクがオンになっていないかだけ注意してください

画面左下でマイクのオンオフの切り替え

21



ここでマイクの
オンオフを切り替え



マイクがオフの状態