

クレジット:

UTokyo Online Education データマイニング入門 2018 森 純一郎

ライセンス:

利用者は、本講義資料を、教育的な目的に限ってページ単位で利用することができます。特に記載のない限り、本講義資料はページ単位でクリエイティブ・コモンズ 表示-非営利-改変禁止 ライセンスの下に提供されています。

<http://creativecommons.org/licenses/by-nc-nd/4.0/>

本講義資料内には、東京大学が第三者より許諾を得て利用している画像等や、各種ライセンスによって提供されている画像等が含まれています。個々の画像等を本講義資料から切り離して利用することはできません。個々の画像等の利用については、それぞれの権利者の定めるところに従ってください。



授業中で配布した.ipynbファイルでは、課題を完成させるためのコーディングのガイドおよび動作確認用の関数が含まれていますが、本ファイルでは省略されています。

課題1 データの基本的な記述統計の計算

データの特徴を指標（統計値）によって記述することを記述統計と呼びます。データ分析の過程で、記述統計は分析対象のデータを観察し、理解するのに用いられます。記述統計により対象データを理解することは、データの前処理（欠損値、外れ値、異常値の処理）の準備にもなります。記述統計および前処理について次回の授業で詳しく説明します。

代表的な記述統計に平均があります。平均はデータの中心傾向を記述するための指標で、データが n 個の数値要素からなる場合、平均は以下のように計算できます。 x_i はデータの要素を表しています。

$$\frac{1}{n} \sum_{i=1}^n x_i$$

以下では、Pythonの基礎で学んだ内容を元にして、数値データの平均に加えて最大値、最小値を求めるPythonコードを実装します。

Q1

0から100までの整数を要素とする任意の長さのリストを引数で入力として受け取り、以下を要素とするリストを返す `avg_max_min` 関数を完成させてください。

- 入力リストの全ての要素の平均値
- 入力リストの全ての要素の中で最大の値
- 入力リストの全ての要素の中で最小の値

```
In [ ]: def avg_max_min(input_list):
```

`avg_max_min` 関数が完成したら、以下のセルを実行して動作を確認してください。

```
In [ ]: avg_max_min([30,39,29,95,70,67,29,56,45,68])
```

Q2

以下のような形式の"exam_score.csv"ファイルを読み込み、教科名をキー、キーの教科の平均値、最大値、最小値のリストを値とする辞書を返す exam_stat 関数を完成させてください。その際、先に作成した avg_max_min 関数を利用してください。

```
## exam_score.csvファイル
kokugo, shakai, sugaku, rika
30, 43, 51, 63
39, 21, 49, 56
...
```

```
In [ ]: import csv

def exam_stat():
```

exam_stat 関数が完成したら、以下のセルを実行して動作を確認してください。

```
In [ ]: exam_stat()
```