

情報管理

第2回：ライブラリとグラフの描画

今回の講義内容

前回に引き続き、Pythonの基本部分が主な内容となります。

今回は、データを読み込んでグラフを描画する方法について説明します。

- データ解析に必要なライブラリの紹介
 - Pandas : データの読み込みや抽出を行う。
 - Numpy : 数値演算を行う。
 - Matplotlib : グラフを描画する。
- 上記ライブラリを使ったグラフの描画方法

Pythonの基本説明は今回までです。

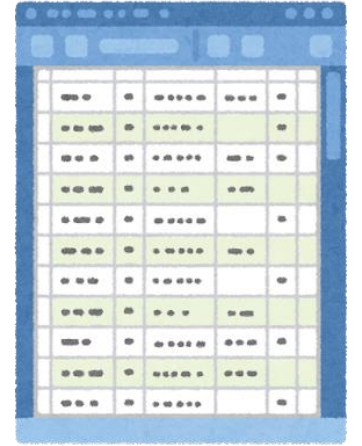
次回以降はPythonを使って具体的にデータ解析の方法を学んでもらいます。

早いうちに基本的な使い方に慣れておきましょう。

データ解析の流れ

主に以下の流れでデータを解析します。

- データを読み込む。
- 読み込んだデータをもとに何らかの処理を行う。
- 処理結果をもとにグラフなどを描画する。

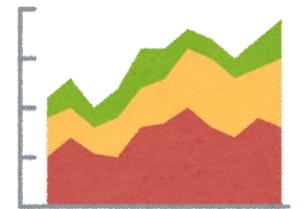
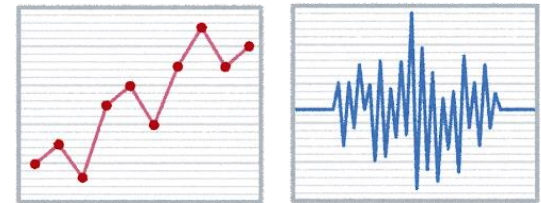


...
...
...
...
...
...
...
...
...
...



$$\hat{x}_{k|k-1} = A\hat{x}_{k-1} + bu_{k-1}$$

$$\hat{x}_k = \hat{x}_{k|k-1} + g_k(y_k - c\hat{x}_{k|k-1})$$

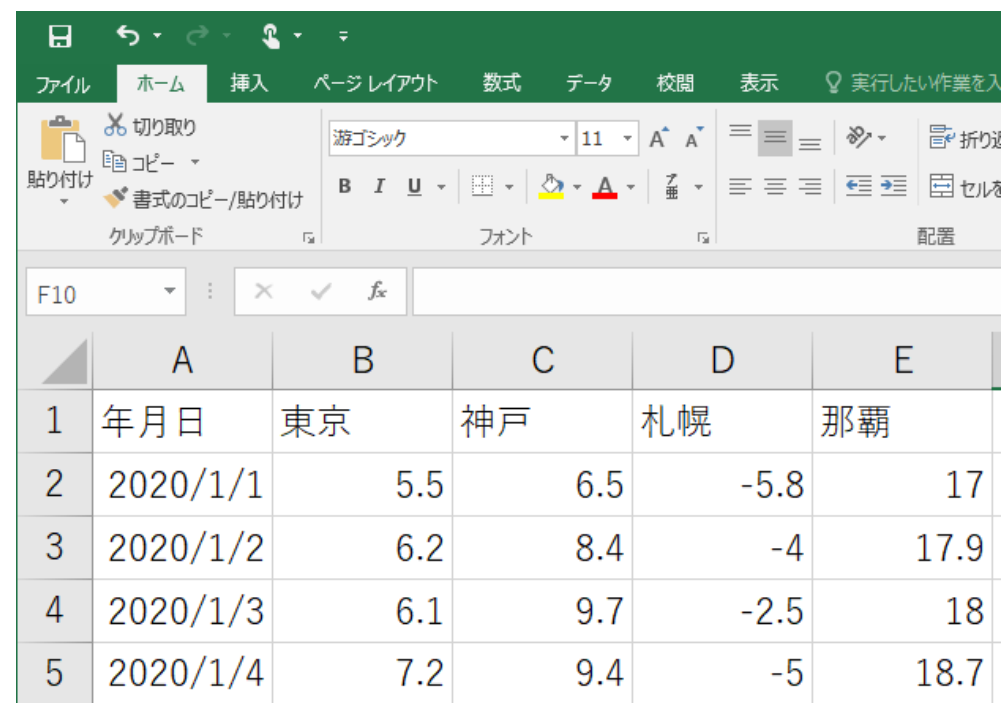


本講義で用いるデータの形式

CSVデータ (Comma-Separated Values Data)

各列の値が「コンマ（,）」で区切られているテキストファイル。

Excelで開くことが可能。



	A	B	C	D	E
1	年月日	東京	神戸	札幌	那覇
2	2020/1/1	5.5	6.5	-5.8	17
3	2020/1/2	6.2	8.4	-4	17.9
4	2020/1/3	6.1	9.7	-2.5	18
5	2020/1/4	7.2	9.4	-5	18.7

kishou - メモ帳

ファイル(E) 編集(E) 書式(O) 表示(V) ヘルプ(H)

年月日, 東京, 神戸, 札幌, 那覇
2020/1/1, 5.5, 6.5, -5.8, 17
2020/1/2, 6.2, 8.4, -4, 17.9
2020/1/3, 6.1, 9.7, -2.5, 18
2020/1/4, 7.2, 9.4, -5, 18.7

テキストエディタで開いた場合

エクセルで開いた場合

Pythonライブラリについて

C言語でMath.hなどのライブラリがあったように，Python にもライブラリが存在します。

Python は豊富なライブラリを利用することで，複雑な計算が簡単に記述することができます。

本講義では，以下 3 種類のライブラリを主に使用します。

- データを読み込む：**Pandas ライブラリ**
- 読み込んだデータをもとに何らかの処理を行う：**Numpy ライブラリ**
- 処理結果をもとにグラフなどを描画する：**Matplotlib ライブラリ**

ライブラリを用いてグラフを描画しよう

ここから先は、実際に Notebook を動かしながら学んでいきましょう。
今回は以下 2 個の Notebook ファイルを動かしてもらいます。

- 02_01_graph.ipynb
3個のライブラリを使って、csvデータを読み込み、簡単な統計値の計算、様々なグラフの描画を行います。
- 02_02_numpy.ipynb
数値演算、特にベクトル・行列演算が得意なライブラリ Numpy について説明しています。

おわりに

データ解析の基礎である，データ読み込み・数値計算・グラフ描画について説明しました。

Python の基本説明は今回までとなります。

次回から，いよいよデータ解析の技術について本格的に解説していきます。

レポート課題

report02_input.csv というデータを読み込み，必要に応じてグラフの描画や計算を行い，データの傾向や性質について考察せよ。

- 計算結果やグラフ描画結果，および考察内容をテキスト部分に記述して提出してください。
- 作成した Notebook (.ipynb) ファイルを，convert_report.ipynb を用いてHTML形式に変換して提出してください。

レポート提出期限：4/26(火) AM10:30

レポート課題 提出例

を調べた。 **xxx を計算/描画し, yyyを調べてみた。(調べた内容の説明)**

(xxx を計算/描画するソースコード)

(計算/描画結果の表示)

**上記の結果から, このデータにはzzzという性質があると
考えられる (結果から導き出される考察)。**

- を見ると,
- がある
 - がある

ことがわかる。このことから, このテストの ことができる。

レポート課題 提出例

考察内容が合っているか否かより、**考察内容が説得力のあるグラフや数値を用いて分かりやすく説明されているか**を重視して評価します。

- 考察内容の根拠としてふさわしいグラフ・数値になっているか。
- グラフの軸ラベルが無かったり、分かりにくい図になっていないか。
- 何を計算しようとしているのか、何を表示しようとしているのかが説明されているか。
(例えば、単に `np.mean(data)` というコードだけを書くのではなく、「データの平均値を調べてみた」というような説明を記述してください。)
- 考察内容は複数あっても良いです。数に応じて加点します。

来週以降のレポート課題も、プログラムだけではなく自分の考えを書いてもらう内容となっています。

今週の課題を通じて、Notebook を使ったレポートの書き方を身につけましょう。