

環境情報論第8回

回帰係数と相関係数

神山 翼, @t_kohyama,
tsubasa@is.ocha.ac.jp,

理3-703

今日は、2つのデータ x と y に
共通する変動を定量化します

回帰係数と相関係数

回帰係数 a : x が動いたとき y がどのくらい動くか

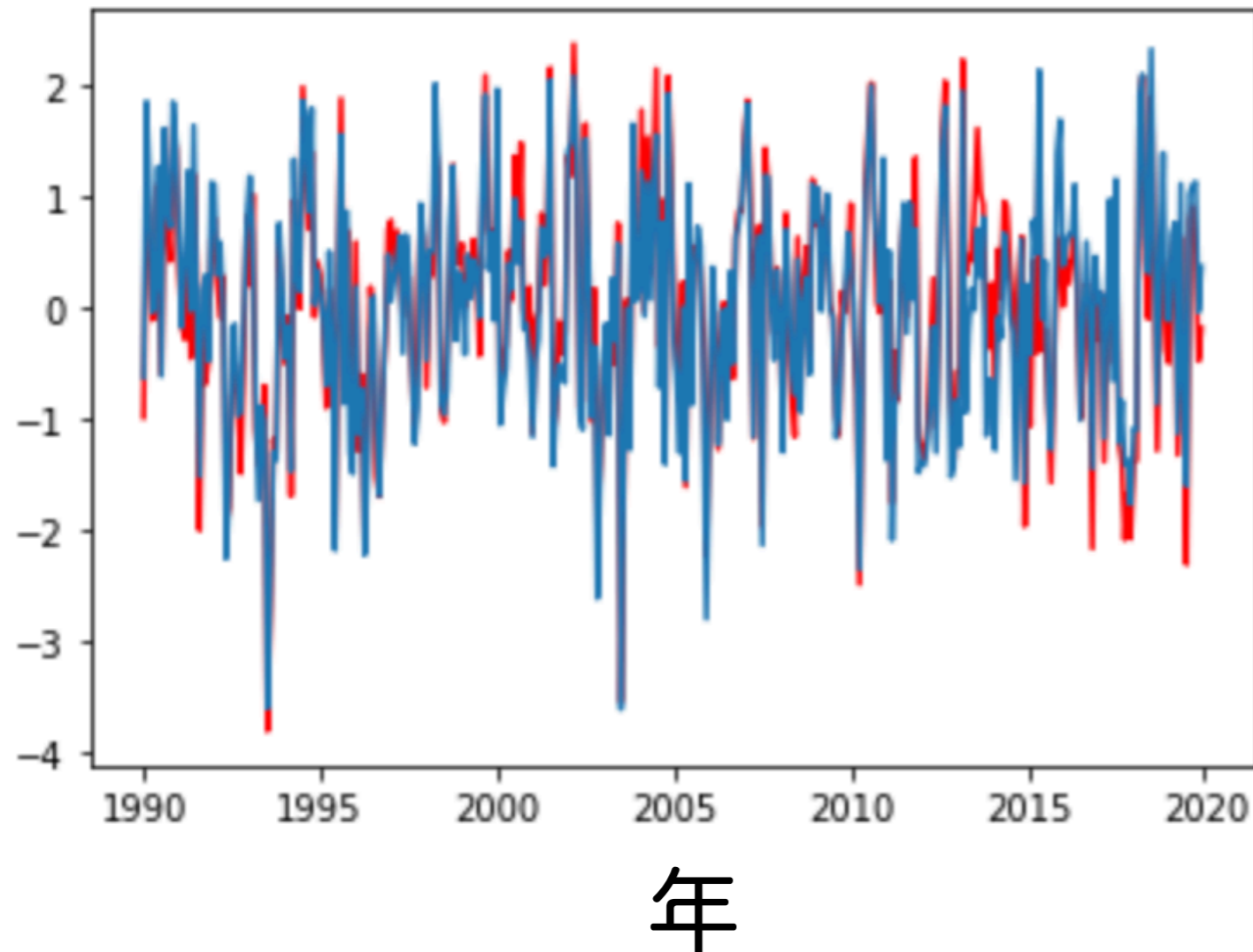
相関係数 r : x と y がどの程度 $y=ax$ に乗っているか

2つのデータ x , y の関係を直線 $y=ax$ で近似して
考察できるようになるのが目標

東京が暖冬だと宇都宮も暖冬

東京（赤）と宇都宮（青）の気温偏差

※季節変動は除去済み。

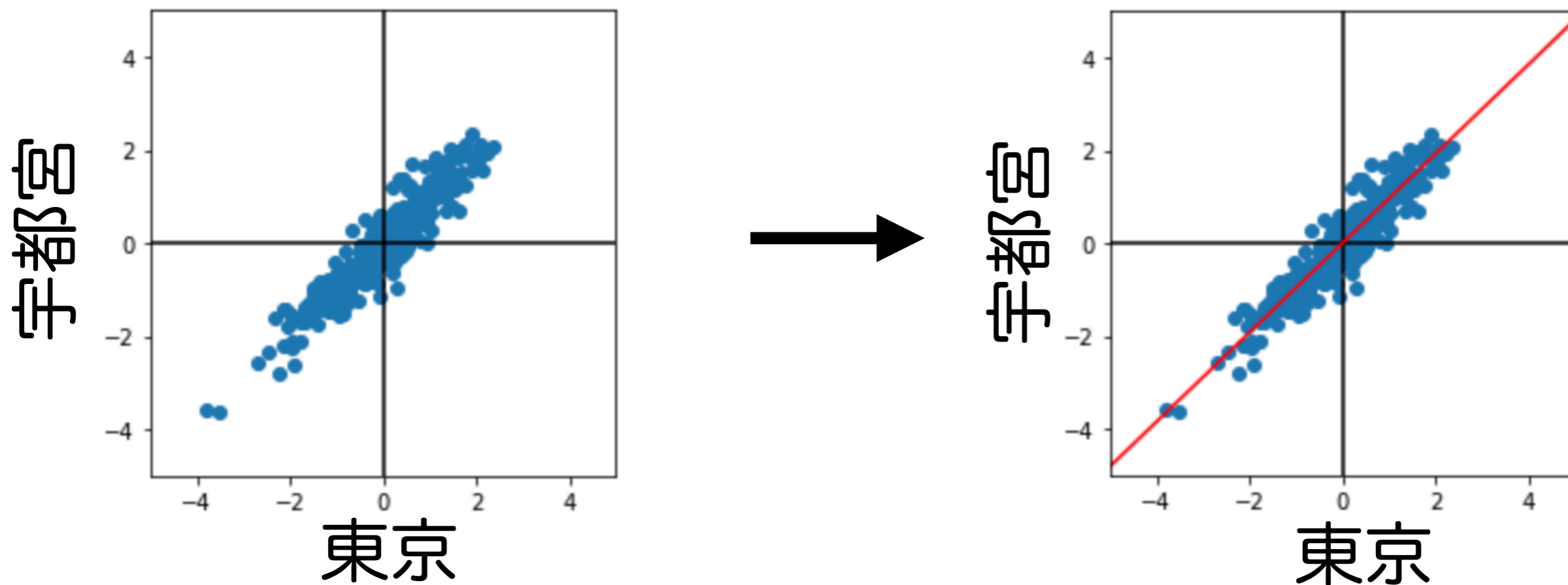


東京が 1°C 上がると、
宇都宮は何 $^{\circ}\text{C}$ 上がる？

東京と宇都宮の気温の
変動は、どのくらい比
例関係にある？

$y=ax$ でエイヤツと近似してみる

「東京が 1°C 上がったとき、宇都宮は何 $^{\circ}\text{C}$ 上がるか」



このときの「直線の傾き a 」が**回帰係数**
単位は $^{\circ}\text{C}/^{\circ}\text{C}$ （あるいは無単位）

回帰係数の計算

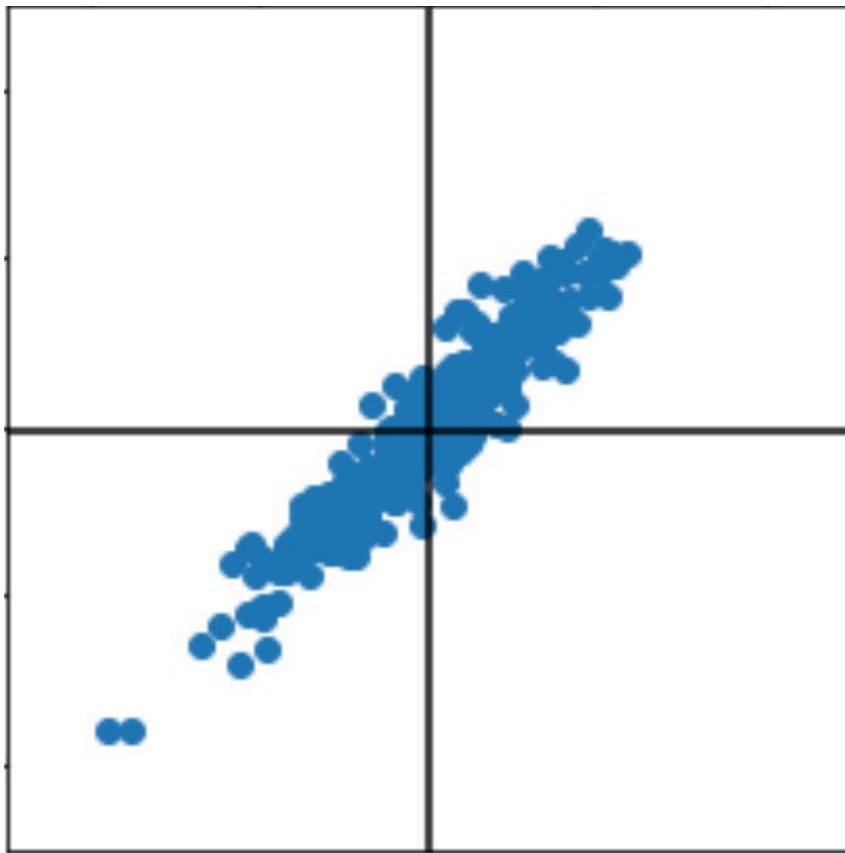
既に出てきた `np.polyfit(x, y, 1)` でOK

```
[a, b] = np.polyfit(tokyoa, utsua, 1)
```

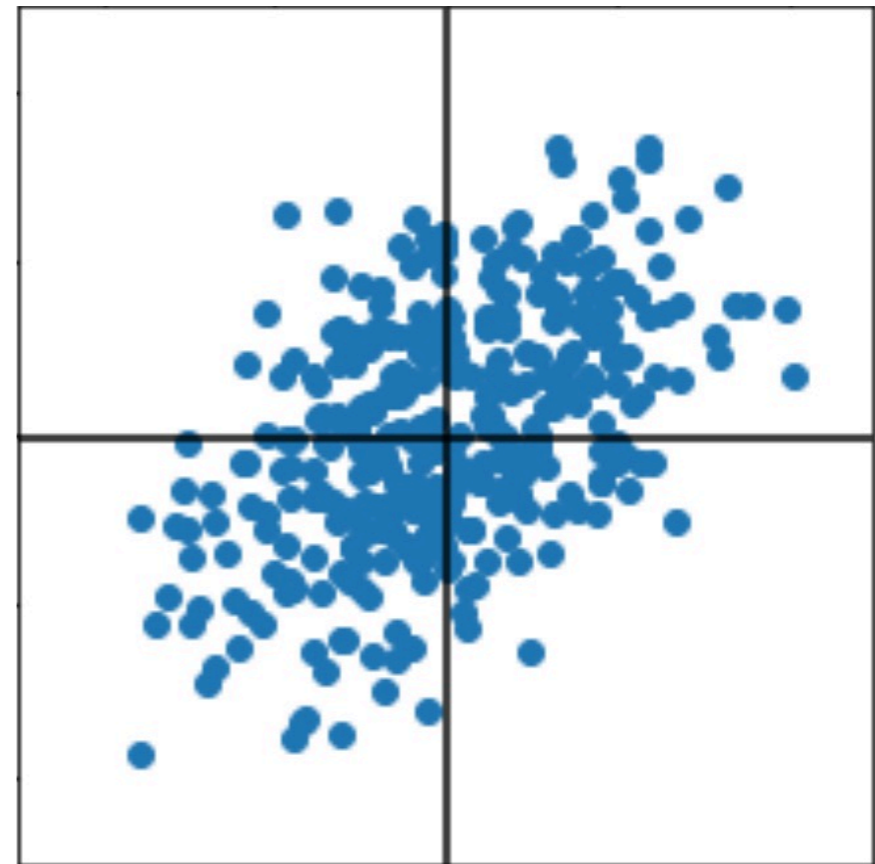
第5回で勉強した「トレンド」は
時刻をxとしたときの回帰係数に他ならない

同じ回帰係数でも，直線への「ノリ」は違うかも

「どれだけちゃんと $y=ax$ に乗っているか」を**相関係数**という



「相関が大きい」

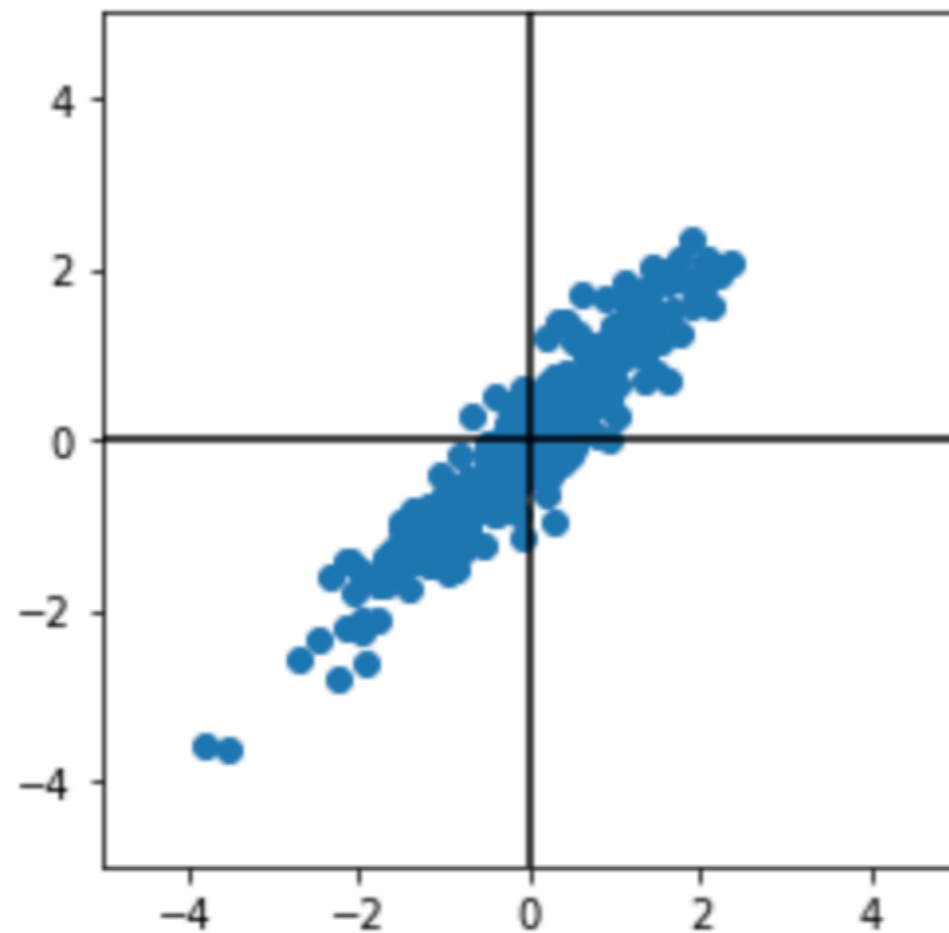


「相関が小さい」

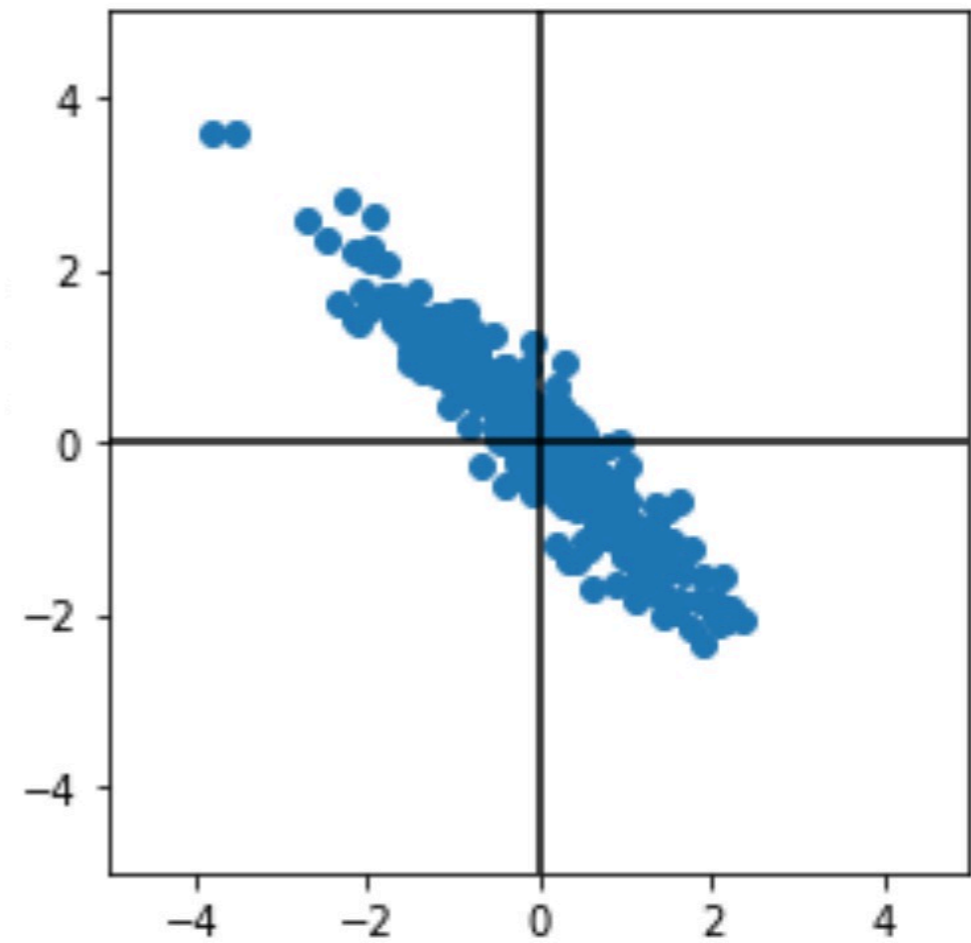
相関係数の符号は， 回帰係数の符号と一致

Aが上がればBも上がる…正の相関

Aが上がるとBは下がる…負の相関



正の相関



負の相関

相関係数の計算

`np.corrcoef(x, y)[1, 0]`と書く

```
np.corrcoef(tokyoa, utsua)[1, 0]
```

```
0.9385239274981371
```

※[1, 0]の意味が知りたい人は「共分散行列」でググってね

相関係数のとりうる範囲は-1から1

絶対値が1に近い相関 = ほぼ完全な比例関係

絶対値が0に近い相関 = 「無相関である」 「直交している」

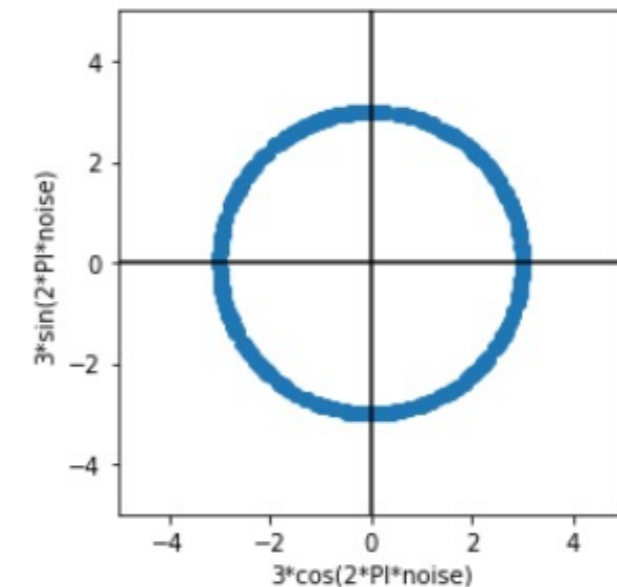
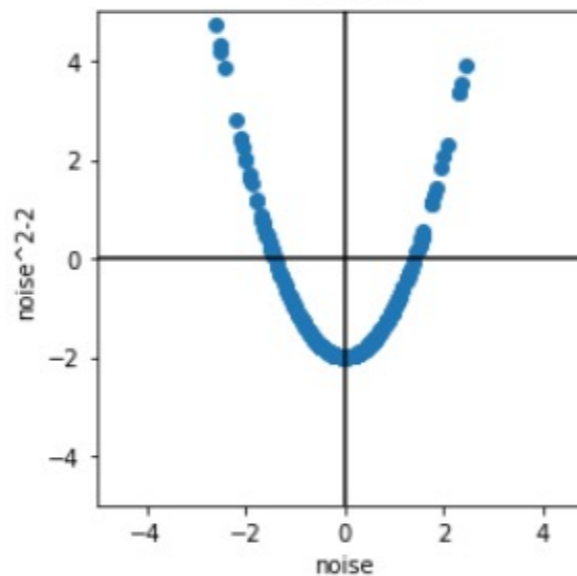
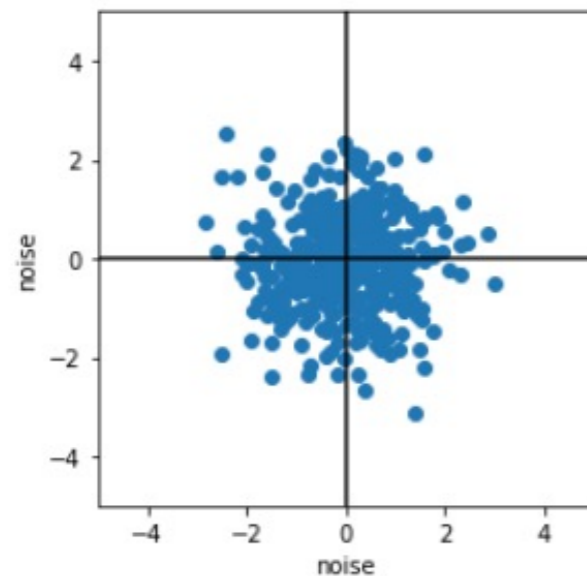
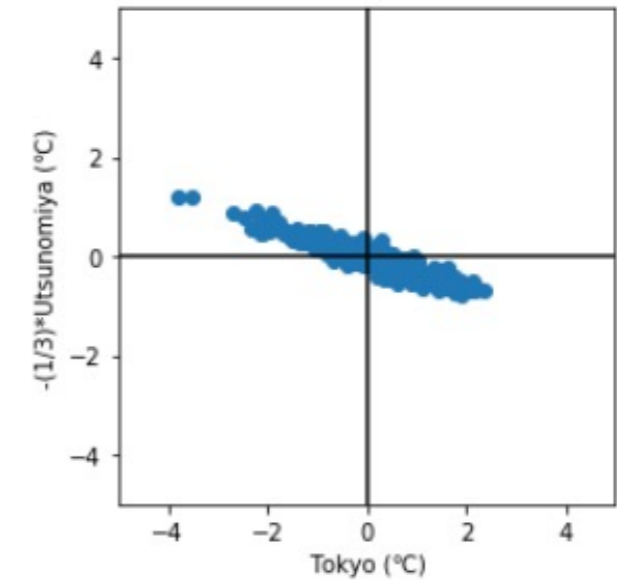
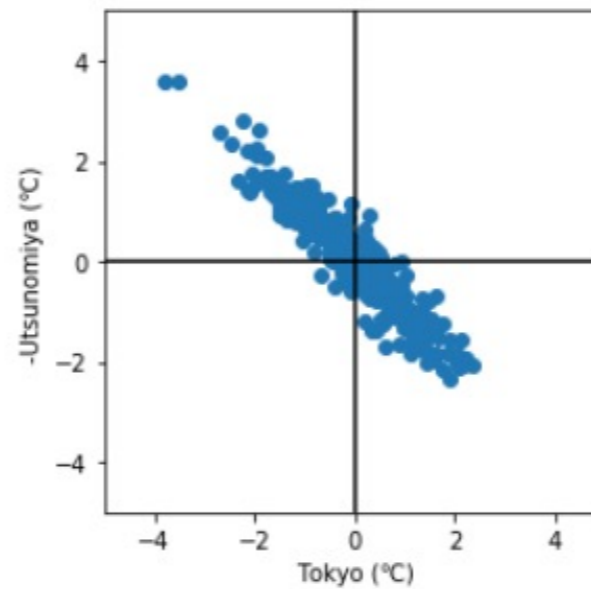
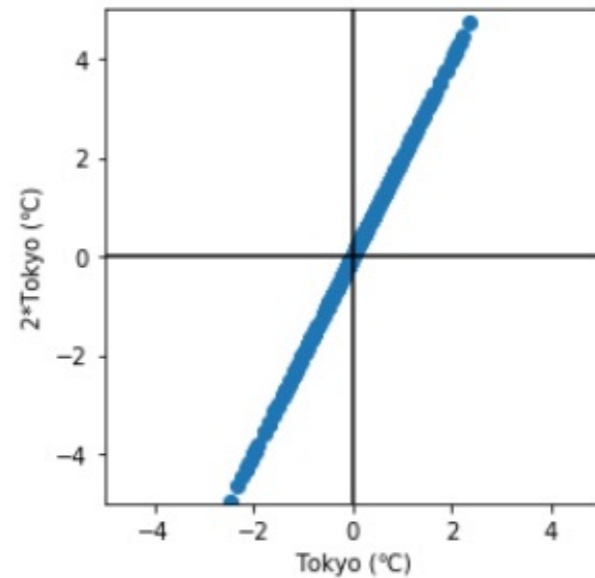
$$r := \cos\theta = \frac{\vec{x} \cdot \vec{y}}{|\vec{x}| |\vec{y}|}$$

相関係数の定義：

データを並べたN次元ベクトルのなす角のコサイン

資料に色々な散布図の回帰係数・相関係数の例を 挙げたので読んでおいてね

相関係数が低いのはどれ？



たのしい課題たち

A: 回帰係数・相関係数の説明

B: 回帰係数・相関係数の身近な例

C: 好きなデータで回帰係数・相関係数の計算

D: 生データの相関と偏差の相関

今日は、2つのデータ x と y に
共通する変動を定量化します

回帰係数と相関係数

回帰係数 a : x が動いたとき y がどのくらい動くか

相関係数 r : x と y がどの程度 $y=ax$ に乗っているか

2つのデータ x , y の関係を直線 $y=ax$ で近似して
考察できるようになるのが目標

本日の導入パートは以上です。
何でも良いのでZoomの方に
授業に関係のあるコメントを
してください（出席代わり）。

コメント拾いが終わったら、
早速今日のプログラミングに進みましょう。