

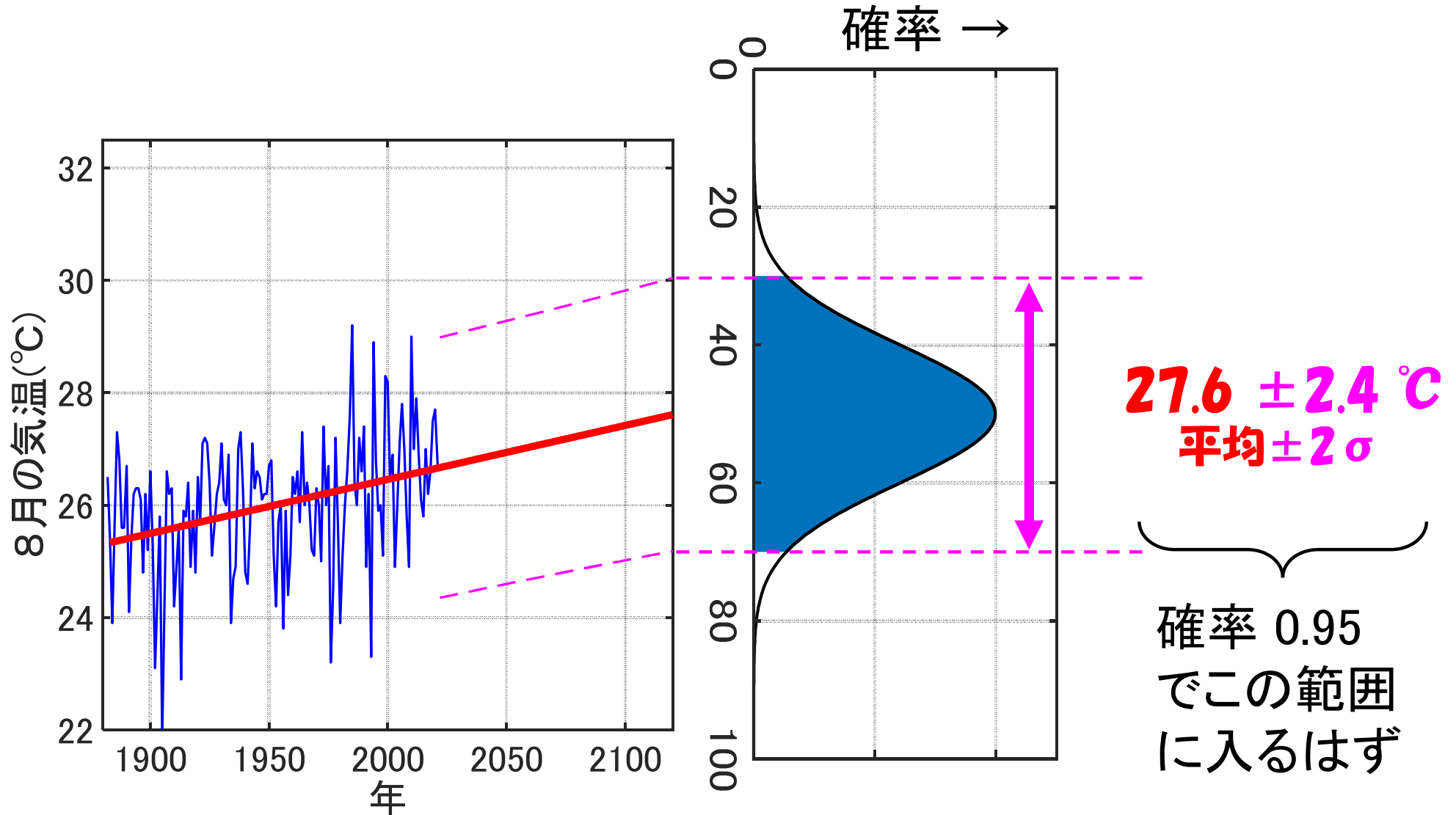
線形回帰で 未来を予測する

その1

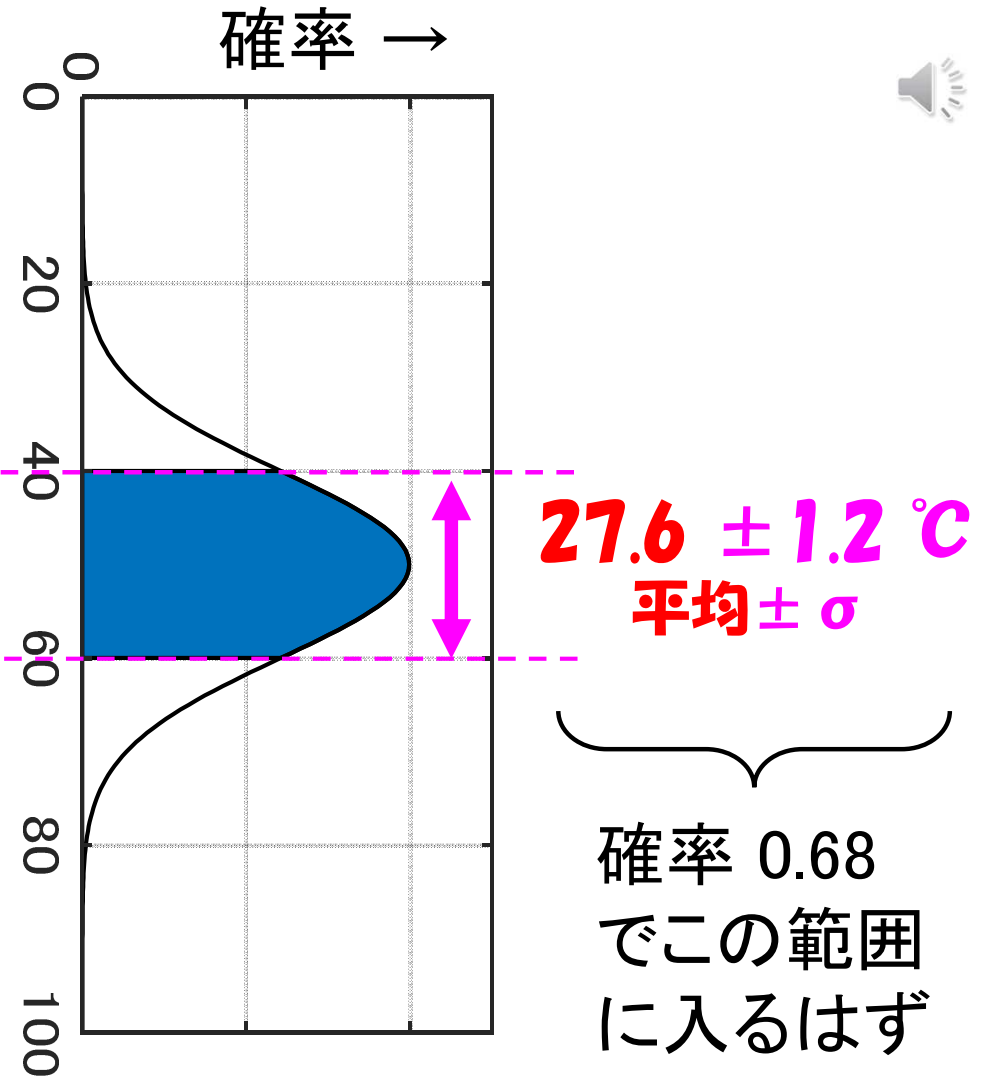
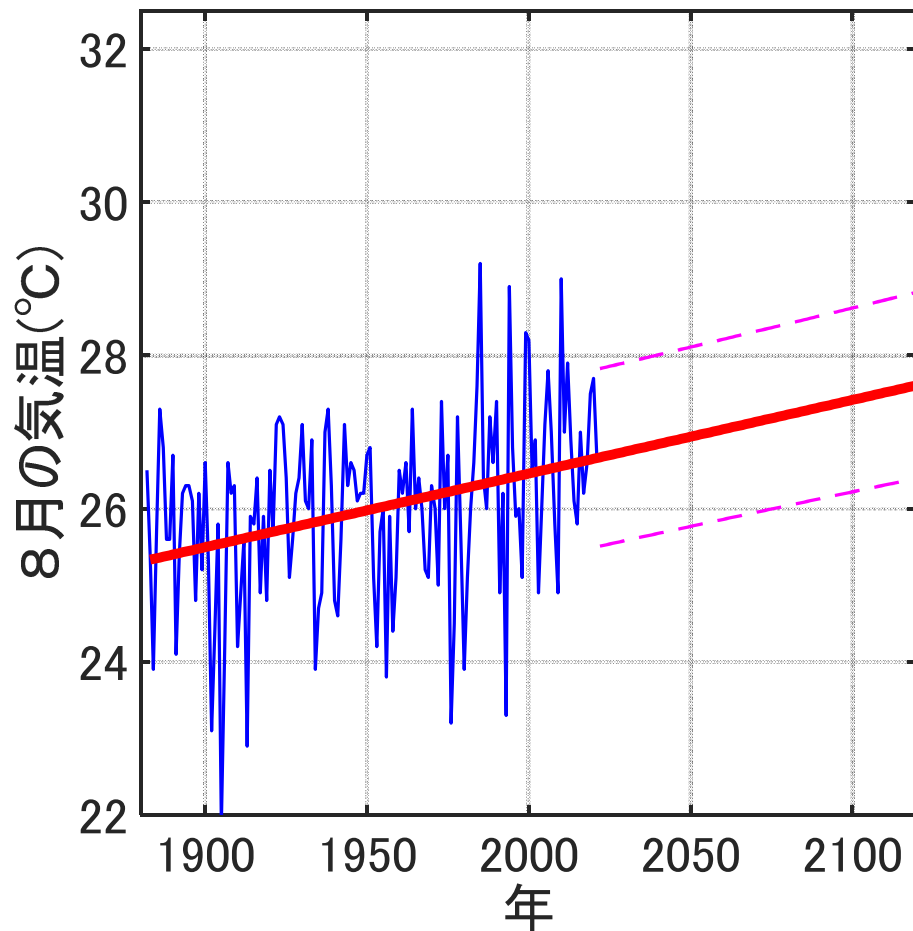
- 身近なデータを可視化
 - データのバラツキを測る
 - 気温、どこまで上がる？
 - その予測は当たるのか？
- 気温のデータを入手
分散と標準偏差
線形回帰で未来予測
予測区間、信頼度



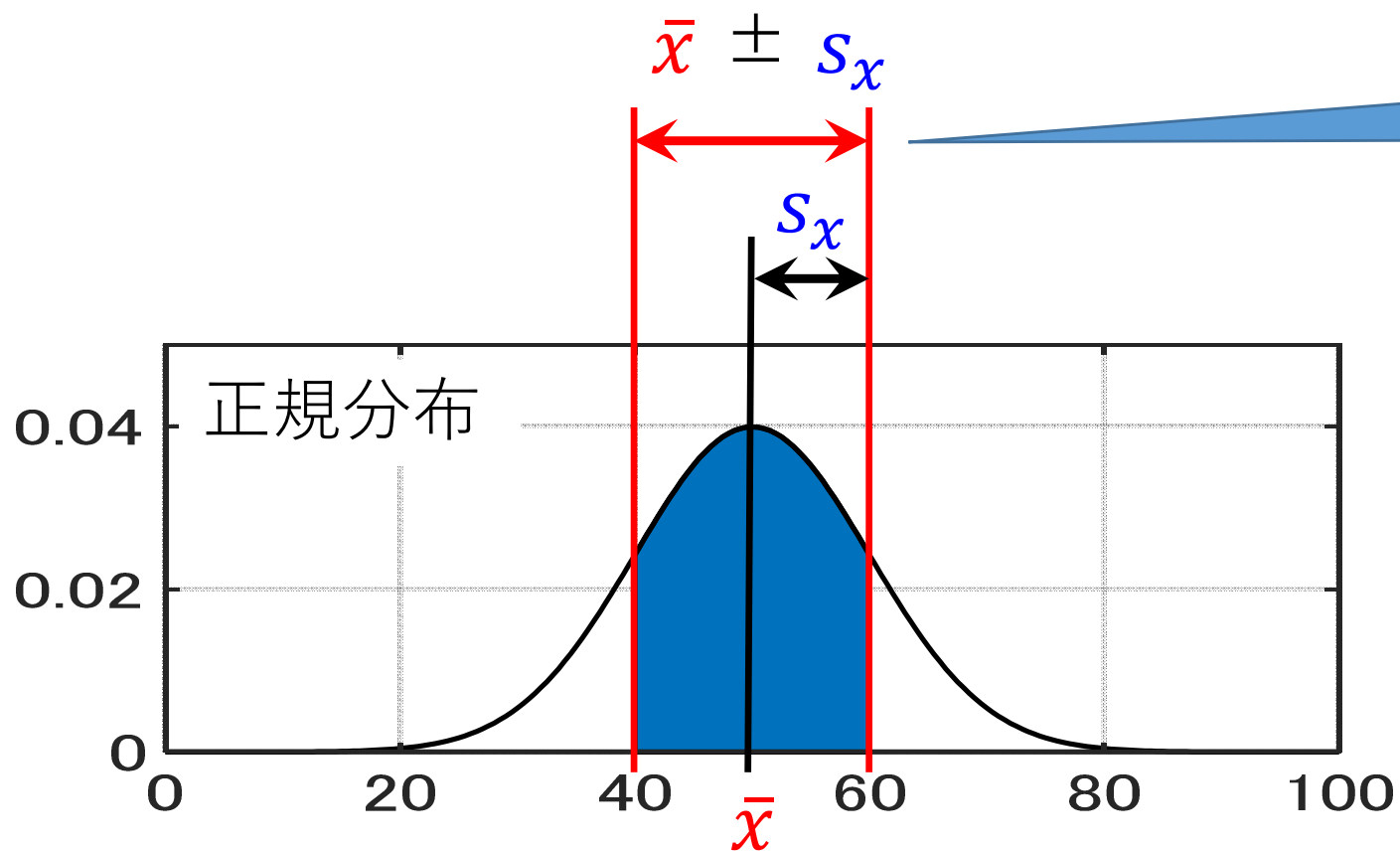
95% の確率でこの範囲内



68% の確率でこの範囲内



正規分布 と 標準偏差



この範囲に全体の
の68.3%が入る

$$\int_{\bar{x}-s_x}^{\bar{x}+s_x} f(x) dx$$

$$= 0.683$$

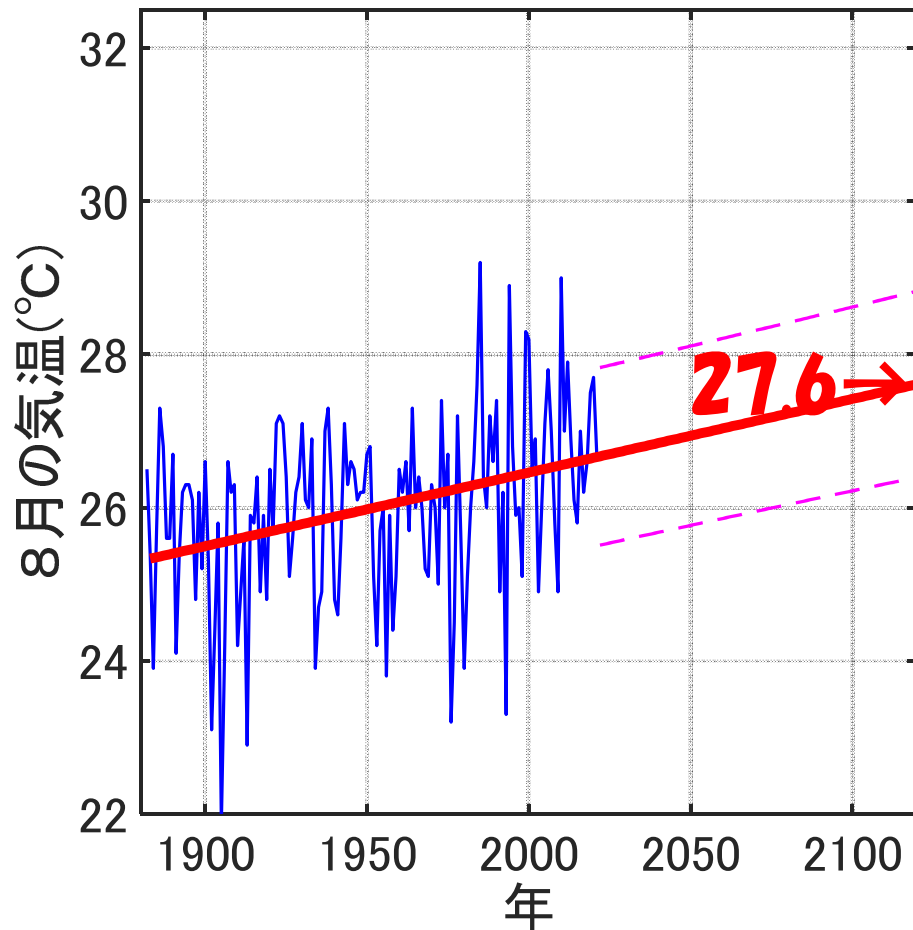
$$f(x) = \frac{1}{\sqrt{2\pi s_x^2}} \exp\left(-\frac{(x - \bar{x})^2}{2s_x^2}\right)$$

\bar{x} : 平均

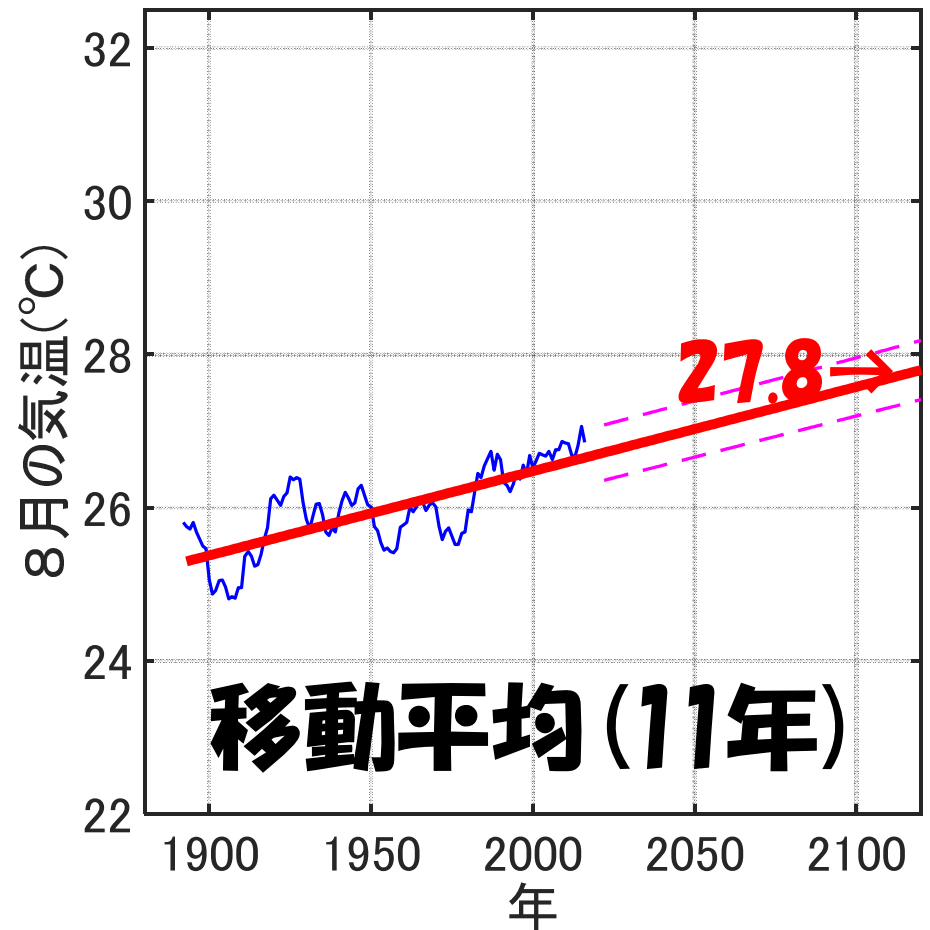
s_x : 標準偏差

移動平均で 予測区間を絞る (直線)

27.6 ± 1.2 °C



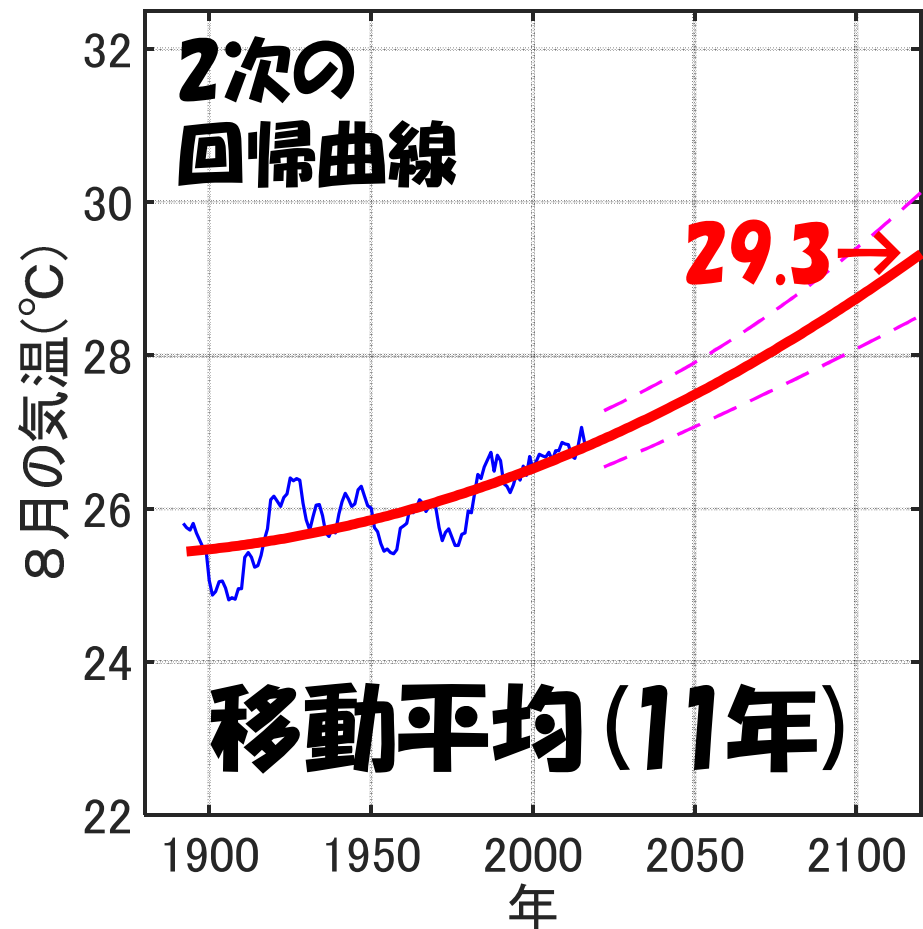
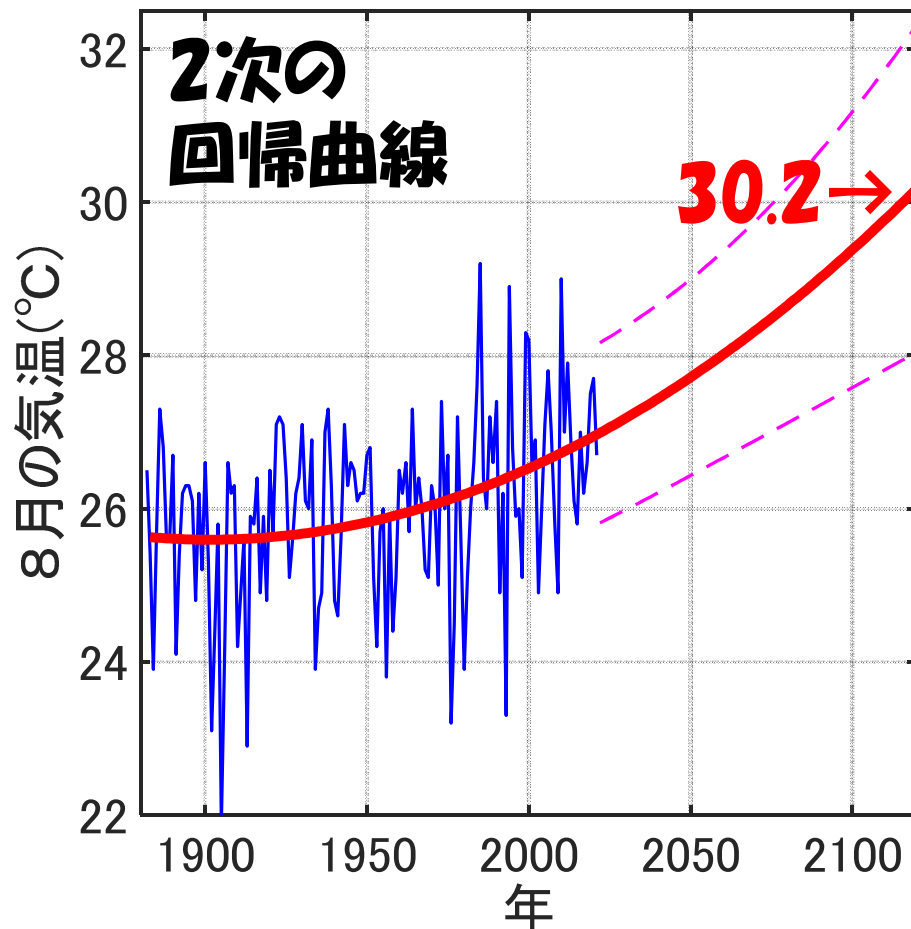
27.8 ± 0.4 °C



移動平均で 予測区間を絞る (曲線)

30.2 $\pm \sigma_k$
(**±大きい**)

29.3 $\pm \sigma_k$
(**±小さい**)



標準偏差と予測区間 (信頼度95%)

$$\hat{y}_k \pm t_{2.5, n-2} \sqrt{\left\{ 1 + \frac{1}{n} + \frac{(x_k - \bar{x})^2}{n s_x^2} \right\} s_r^2}$$

予測値 ±ばらつく幅

回帰の誤差が大きいとばらつく幅も大きい

$$\hat{y}_k = \hat{a} + \hat{b} x_k$$

$$\hat{a} = \bar{y} - \hat{b} \bar{x}$$

$$\hat{b} = \frac{s_{xy}}{s_x^2}$$

$$s_r^2 = \frac{1}{n-2} \sum_{i=1}^n (y_i - \hat{y}_i)^2$$

回帰の誤差の大きさ

$$\left\{ \begin{array}{l} \bar{x} = \frac{1}{n} \sum_{i=1}^n x_i \\ \bar{y} = \frac{1}{n} \sum_{i=1}^n y_i \\ s_x^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2 \\ s_{xy} = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y}) \end{array} \right.$$

MATLAB

エクセルファイルから
データを読み込む

データを入力すると
回帰多項式

$$p_0 + p_1x + p_2x^2 + \dots + p_nx^n$$

の係数 p_n を出力する

予測の値とその信頼度
(予測区間)を出力する

```
clear variables; close all;
D=readmatrix('長岡の気温(月).xls');
%-----
figure('Position',[10 10 300 300]);
figure(1); % 全てのデータ
%-----
X=D(:,4); % 3:長岡市、4:新潟市
[N1,N2]=size(X);
%----- 気温のデータ
x=[0:N1-1]'; x=x/12; x=x+1882;
plot(x,X); axis('tight'); grid on;
xlabel('年'); ylabel('月平均気温(°C)');
%----- 各年の平均
Y=reshape(X,12,N1/12);
y=mean(Y)';
x=[0:N1/12-1]'; x=x+1882;
plot(x,y); axis('tight'); grid on;
xlabel('年'); ylabel('気温(°C)');
axis([1882 2120 10 20]);
%----- 線形回帰
n=1;


[p,S]=polyfit(x,y,n);


x=[0:N1/12-1+100]'; x=x+1882;


[yy,delta]=polyval(p,x,S);


hold on; plot(x,yy,'r'); % 予測値
x=[N1/12:N1/12+100]'; x=x+1882;
[yy,delta]=polyval(p,x,S);
hold on; plot(x,yy+2*delta,'m--');
hold on; plot(x,yy-2*delta,'m--');
```


初版： 2022年7月

制作： 岩橋政宏

所属： 長岡技術科学大学