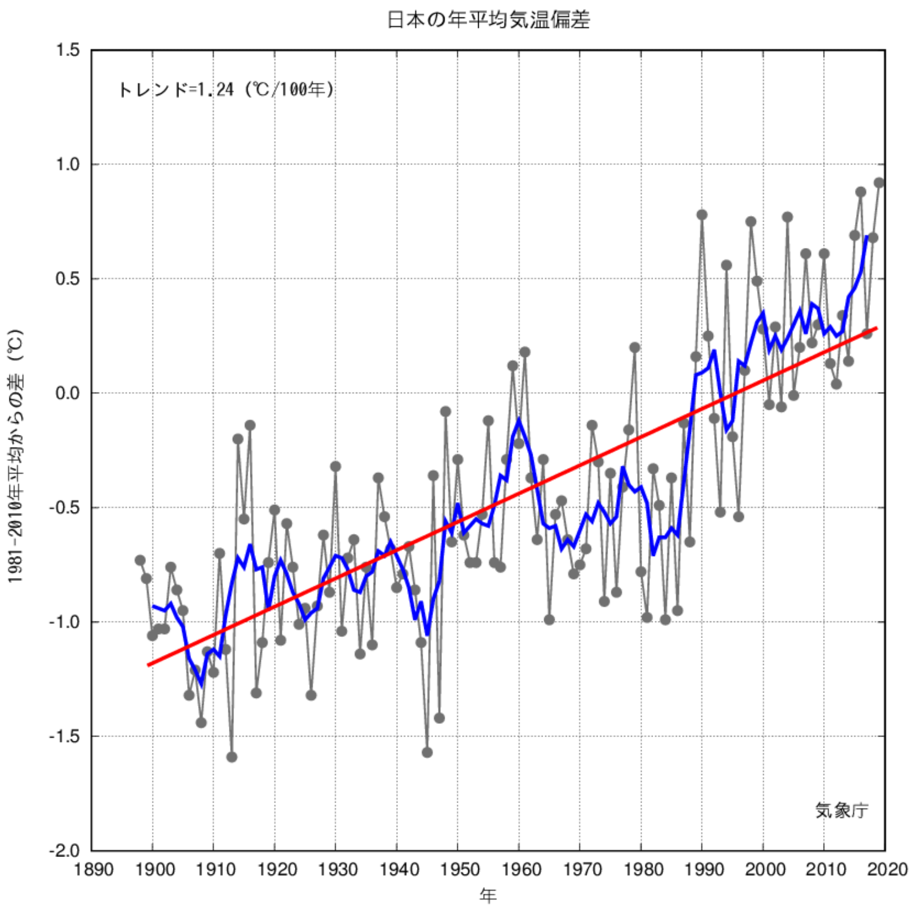
いまさら Fortran 入門

はじめに

- 今後の実験や実習では，取得したデータを様々な方法で処理することになるだろう．
- これまでに学んだ命令を組み合わせてデータ解析しよう．
  - 平均，分散，標準偏差，共分散の計算
  - 移動平均（右図の青線）
  - 最小二乗法による直線フィッティング（例えば右図の赤線）



http://www.data.jma.go.jp/cpdinfo/temp/an_jpn.html
2020/01/07 にダウンロードしました.

復習：平均, 分散, 標準偏差

- 下のようなデータがあるとする.

$$(x_1, y_1), (x_2, y_2), \dots, (x_N, y_N)$$

（x_i, y_i は例えば前ページの時間と気温）

  - 平均 (average, mean) $\bar{y} = \frac{1}{N}\sum_{i=1}^{N} y_i$
  - 分散 (variance) $\sigma_y^2 = \frac{1}{N}\sum_{i=1}^{N} (y_i - \bar{y})^2$
    - 標準偏差 (standard deviation) は σ_y

復習: 共分散, 相関係数

- 下のようなデータがあるとする.

$$(x_1, y_1), (x_2, y_2), \ldots, (x_N, y_N)$$

$$(x_1, z_1), (x_2, z_2), \ldots, (x_N, z_N)$$

（x_i, y_i, z_i は例えば時間, 気温, 降水量）

  - 共分散 (covariance) $C_{yz} = \frac{1}{N}\sum_{i=1}^{N}(y_i - \bar{y})(z_i - \bar{z})$
  - 相関係数 (correlation coefficient) $r = \frac{C_{yz}}{\sigma_y \sigma_z}$

移動平均

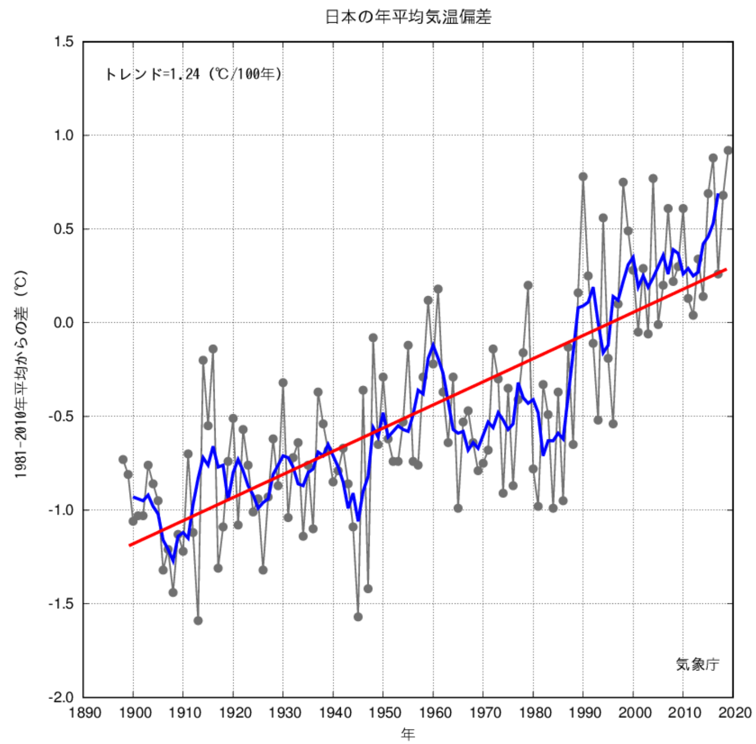
- 下のようなデータがあるとする.

$$(x_1, y_1), (x_2, y_1), \ldots, (x_N, y_N)$$

- 大規模な変化のみを取り出す方法として, しばしば移動平均が用いられる.
  - 移動平均では, ある区間ごとに平均を求める.

$$(\bar{x}_1, \bar{y}_1), (\bar{x}_2, \bar{y}_2), \ldots, (\bar{x}_{N-n+1}, \bar{y}_{N-n+1})$$

$$\bar{x}_j = \frac{1}{n} \sum_{i=j}^{j+n-1} x_i \qquad \bar{y}_j = \frac{1}{n} \sum_{i=j}^{j+n-1} y_i$$



http://www.data.jma.go.jp/cpdinfo/temp/an_jpn.html
2020/01/07 にダウンロードしました.

青線は 5 年間の移動平均の結果

最小二乗法
概要 (1)

- 下のようなデータがあるとする.

$$(x_1, y_1), (x_2, y_1), \dots, (x_N, y_N)$$

  – そして, これらの各データ, y_i, の分散（誤差）が下のように書けるとする.

$$\sigma_1, \sigma_2, \dots, \sigma_N$$

- このとき, このデータに対して, 下の関数をフィッティングすることを考えよう.

$$y = f(x)$$

最小二乗法
概要 (2)

- 「フィッティングする」とは, 言うなれば「よく合っているようにする」ことである.
- 「よく合っている」の基準の一つは, 二乗誤差が小さいことである.
  - つまり, 下の値が最小となれば「よく合っている」だろう.

$$E = \sum_{i=1}^{N} \left(\frac{y_i - f(x_i)}{\sigma_i} \right)^2$$

- 二乗誤差が極小となるのは, その微分がゼロになるときである. この条件を基に, $y = f(x)$ を求めればよい.

最小二乗法
一次関数 (1)

- ここから先は, 具体的な例を使って考える.
- フィッティングしたい関数が一次関数の時を考える.

$$y = ax + b$$

- 二乗誤差は下のようになる.

$$E = E(a, b) = \sum_{i=1}^{N} \left(\frac{y_i - ax_i - b}{\sigma_i} \right)^2$$

- そして, 一次関数の係数を, 下の条件を基に決めればよい.

$$\frac{\partial E(a, b)}{\partial a} = 0 \qquad \frac{\partial E(a, b)}{\partial b} = 0$$

最小二乗法
一次関数 (2)

- 先の式より, 下のような, a, b の連立方程式となる.

$$\frac{\partial E(a,b)}{\partial a} = -\sum_{i=1}^{N} \frac{2x_i}{\sigma_i}\left(\frac{y_i - ax_i - b}{\sigma_i}\right)$$

$$= 2a\sum_{i=1}^{N}\frac{{x_i}^2}{{\sigma_i}^2} + 2b\sum_{i=1}^{N}\frac{x_i}{{\sigma_i}^2} - 2\sum_{i=1}^{N}\frac{x_i y_i}{{\sigma_i}^2} = 0$$

$$\frac{\partial E(a,b)}{\partial b} = -\sum_{i=1}^{N} \frac{2}{\sigma_i}\left(\frac{y_i - ax_i - b}{\sigma_i}\right)$$

$$= 2a\sum_{i=1}^{N}\frac{x_i}{{\sigma_i}^2} + 2b\sum_{i=1}^{N}\frac{1}{{\sigma_i}^2} - 2\sum_{i=1}^{N}\frac{y_i}{{\sigma_i}^2} = 0$$

最小二乗法
一次関数 (3)

- 先の式は, a, b に関する連立一次方程式であるから, a, b は下のように求められる.

$$a = \frac{\left(\sum_{i=1}^{N} \frac{x_i y_i}{{\sigma_i}^2}\right)\left(\sum_{i=1}^{N} \frac{1}{{\sigma_i}^2}\right) - \left(\sum_{i=1}^{N} \frac{x_i}{{\sigma_i}^2}\right)\left(\sum_{i=1}^{N} \frac{y_i}{{\sigma_i}^2}\right)}{\left(\sum_{i=1}^{N} \frac{{x_i}^2}{{\sigma_i}^2}\right)\left(\sum_{i=1}^{N} \frac{1}{{\sigma_i}^2}\right) - \left(\sum_{i=1}^{N} \frac{x_i}{{\sigma_i}^2}\right)^2}$$

$$b = \frac{\left(\sum_{i=1}^{N} \frac{{x_i}^2}{{\sigma_i}^2}\right)\left(\sum_{i=1}^{N} \frac{y_i}{{\sigma_i}^2}\right) - \left(\sum_{i=1}^{N} \frac{x_i y_i}{{\sigma_i}^2}\right)\left(\sum_{i=1}^{N} \frac{x_i}{{\sigma_i}^2}\right)}{\left(\sum_{i=1}^{N} \frac{{x_i}^2}{{\sigma_i}^2}\right)\left(\sum_{i=1}^{N} \frac{1}{{\sigma_i}^2}\right) - \left(\sum_{i=1}^{N} \frac{x_i}{{\sigma_i}^2}\right)^2}$$

実習へ

- 実習で自分でプログラムを書き,
  - 平均, 分散, 標準偏差を計算してみよう
  - 移動平均してみよう
  - 最小二乗法を使って一次関数を求めてみよう.