

本日の演習内容

最小二乗法と回帰直線

科学実験ではさまざまな物理量のあいだの関係を検証することが頻繁に行われる。

たとえば、物理量 x と y の関係を調べるには、 x の値を徐々に変化させ、 y の値がどのように変化するかを調べればよい。しかし、実験には誤差がつきものである。実験条件の微妙な違いによって、 x や y の値はばらついてしまう。よって、誤差の影響をできるだけ排除して、 x と y の関係をできるだけ正確に求めるために、**最小二乗法（回帰分析）** という手法が用いられる。

最小二乗法の考え方

n 個の実験データ点のうち、 i 番目のデータ点を (x_i, y_i) とすると、理論曲線 $f(x)$ からのずれ（残差）は、

$$\Delta_i = y_i - f(x_i)$$

と定義できるが、これの和 $\sum_i \Delta_i$ は $i \rightarrow \infty$ でゼロに近づいていく。しかし、残差の自乗 Δ_i^2 の和はゼロにはならない。

$$S = \sum_i \Delta_i^2 = \sum_i [y_i - f(x_i)]^2$$

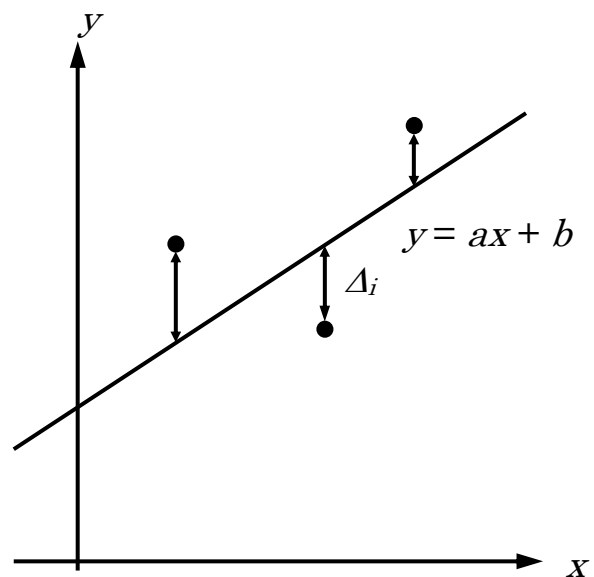
この S を最小にするように $f(x)$ を決めることができれば、理論式により実験値の再現ができたことになる。

一次関数

測定量 y が x と、次式のような比例関係（線形関係）にある簡単な場合（一次関数）について、考えてみよう。

$$y = ax + b$$

ここで、 a は直線の傾き、 b は Y 軸の切片である。一次関数に関して、最小自乗法は解析的に解くことができる。各データ点 (x_i, y_i) は実験から得られるので、 a と b



を変数だと考えて、 S が最小となる a と b の値を求めればよい。つまり、 S の値が最小となるとき、 S の a と b による偏微分は、どちらもゼロとならなければならない。

$$\frac{\partial S}{\partial a} = \frac{\partial S}{\partial b} = 0$$

この式より、

$$\frac{\partial S}{\partial a} = \frac{\partial}{\partial a} \sum_i^n [y_i - (ax_i + b)]^2 = -2 \sum_i^n x_i [y_i - (ax_i + b)] = 0$$

$$\frac{\partial S}{\partial b} = \frac{\partial}{\partial b} \sum_i^n [y_i - (ax_i + b)]^2 = -2 \sum_i^n [y_i - (ax_i + b)] = 0$$

さらに、

$$\sum_i^n x_i y_i - a \sum_i^n x_i^2 - b \sum_i^n x_i = 0$$

$$\sum_i^n y_i - a \sum_i^n x_i - nb = 0$$

となるので、

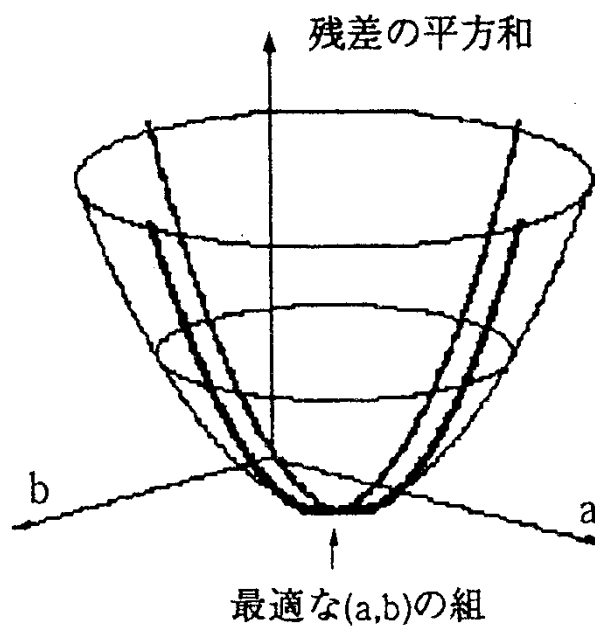
$$b = \frac{1}{n} \sum_i^n y_i - \frac{a}{n} \sum_i^n x_i = \bar{y} - a\bar{x}、\text{ただし、}\bar{y} = \frac{1}{n} \sum_i^n y_i、\bar{x} = \frac{1}{n} \sum_i^n x_i$$

とすると、

$$a = \frac{\sum_i^n x_i y_i - n\bar{x} \cdot \bar{y}}{\sum_i^n x_i^2 - n\bar{x}^2}$$

$$b = \frac{\bar{y} \sum_i^n x_i^2 - \bar{x} \sum_i^n x_i y_i}{\sum_i^n x_i^2 - n\bar{x}^2}$$

がもとまる。こうして得られた直線は x に対する y の**回帰直線**という。



課題 1 回帰直線

次の 7 個のデータの組 $(x_1, y_1) \cdots (x_7, y_7)$ を用いて「散布図」(データポイントのみのもの)を作成し、データポイントをもっともうまく通る直線(データポイントと直線の差をもっとも小さくなるような直線)を引きなさい。

x_i	0	2	4	6	8	10	12
y_i	2.1	5.0	9.0	12.6	17.3	21.0	24.7

- (1) Excel を起動する。
- (2) セル A1 に「 x_i 」、セル B1 に「 y_i 」と入力する(見出しの作成)。
- (3) セル範囲 A2: A8 に「 x_1 から x_7 のデータ」を、セル範囲 B2: B8 に「 y_1 から y_7 のデータ」を入力する。
- (4) 散布図を作成する。
- (5) セル A11 に x の平均値「 $\sum x_i/n$ 」を、セル B11 に y の平均値「 $\sum y_i/n$ 」の計算式を入れる。
(=SUM(A2:A8)/7、=SUM(B2:B8)/7)
- (6) セル C1 に「 x_i^2 」と入力し、セル範囲 C2: C8 に「 x_1^2 から x_7^2 」を計算させ、セル C11 に「 $\sum x_i^2$ 」を計算させる。
- (7) セル D1 に「 $x_i y_i$ 」と入力し、セル範囲 D2: D8 に「 $x_1 y_1$ から $x_7 y_7$ 」を計算させ、セル D11 に「 $\sum x_i y_i$ 」を計算させる。
- (8) セル G2 に「傾き(a)」とラベルし、セル H2 に切片を計算させる。
- (9) セル G3 に「切片(b)」とラベルし、セル H3 に切片を計算させる。
- (10) セル H2 と H3 の傾きと切片を用い、A2:A8 の x データを用い、E2:E8 に回帰分析の結果を計算する。E1 には「結果」とラベルしておく。
- (11) 「結果」を散布図に追加する。(やり方は授業で説明する)

課題 2

サーミスタの温度 $t(^{\circ}\text{C})$ と抵抗 $R(\text{k}\Omega)$ の関係は次のように表される。

$$R = R_0 \exp(A / (t + 273.15))$$

下のデータを用いて、上の関係を最適化する R_0 と B の値を求めよ。

t_i	0	10	20	40	50	60	70	80	90	100
R_i	110.0	70.0	46.0	21.5	15.0	11.0	8.1	6.0	4.6	3.5

(考え方)

- (1) 上式の両辺の対数を取り、 $\ln R_i = A / (t_i + 273.15) + \ln R_0$ とする。
- (2) $y_i = \ln R_i$ 、 $x_i = 1 / (t_i + 273.15)$ とおくと、 $y_i = A \times x_i + \ln R_0$ となる。
- (3) ここで、傾き a は A 、切片 b は $\ln R_0$ である。
- (4) 課題 1 と同様な方法で散布図に回帰直線を挿入したものを作製する。
- (5) R_0 の値を求める。(切片 b から「=EXP(b)」を使う)

- 完成したら、ファイルをメールに添付して提出すること。

送り先 : [ynagasa あつと fc.ritsumei.ac.jp](mailto:ynagasa@fc.ritsumei.ac.jp)

- メールタイトルは” 基礎演習 2 レポート : 氏名と学生証番号” とすること。
- 提出ファイルの最初のページの右上には必ず 氏名と学生証番号を記入 すること。
- ファイル名は “氏名+171018” とする。(例 : 立命太郎 171018)
(「.xlsx」等の拡張子は自動で付くので、自分で書き込む必要はない)
- ファイル名の「氏名」以外は必ず半角。
- 最終締め切り : 2017 年 10 月 24 日 (火曜)。
- 日常点評価の提出課題は定期試験と同等の意味があるので真剣に取り組むこと。