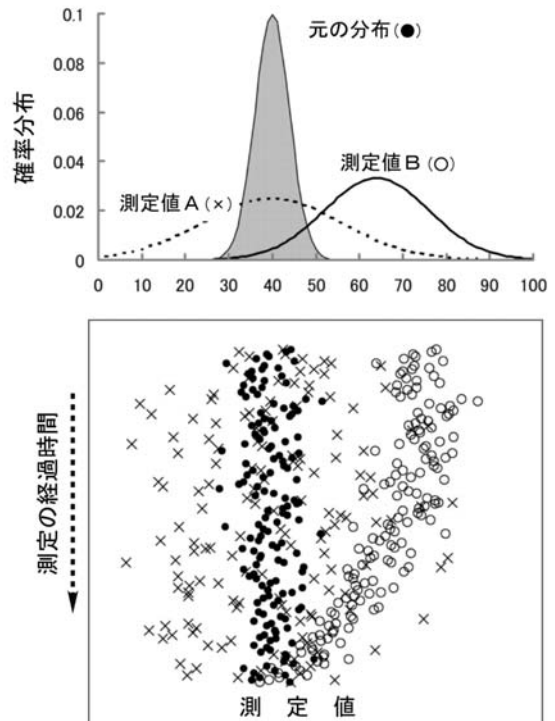


測定と誤差について

実習では、いろいろな定量化の方法を学ぶ。個体や細胞の大きさや重量、酵素反応で生成した物質の量や濃度、DNA やポリペプチドのゲル電気泳動で移動した距離や分離して得た物質の量など、専用の測定機器を使って数値化する。個体数や行動パターンのように観察者が1つ1つカウントすることによって数値化する場合もあるだろう。実験の種目ごとに定量化の手法はさまざま異なるが、その後のデータ処理では、統計学的には同じ方法を用いる。数値データの取り扱いについて、一般的なルールをここでは解説する。

§ 平均値と標準偏差

右の図は、ある実験で、得られるデータの分布を示す。●（灰色で塗ったピーク）は、生物試料の元の分布を示す。重さや長さなど、数値化できるパラメータの本来の分布^{注1}である。それに合わせて、実験者が計測して求めた値の例を2つ、測定値A（破線、×）と測定値B（実線、○）を示す。このように値のばらつきの程度を示すグラフを確率分布とよぶ。図の下には、同じ計測実験を繰り返したときの1つ1つの計測値の変動（縦方向は経過時間）を示す。



私たちの身長が一人一人違うように、生体試料の特性を表した元の値（ここでは x と書く）は、けっして同じではない。個体差や実験条件などが原因となってばらつき、必ずある広がりを持った分布（灰色で示された分布）となる。一般に、この分布は、正規分布（ガウス分布）、

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \cdot \exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right] \quad \dots 1)$$

の形で表記することができる。この式の中の μ を平均値^{注1}（mean value、上の例では $\mu = 40$ ）、 σ を標準偏差^{注1}（S.D.、standard deviation、上の例では $\sigma = 4$ ）という。標準偏差は、元のばらつきの大きさ、分布の広がりを示すパラメータである。もし、私たちの測定実験の精度が非常に高く、いつも正確な値を得られると仮定すると、測定値は図中の●印のデータが示すような、元の分布と同じばらつきになると考えられる。測定回数を多くすればするほど、計測した結果は、ますます、正しい元の正規分布の様子を正確に反映するようになるであろう。データ数が多くなると、求める平均値や標準偏

差が、さらに正確になる。一般に、

$$m = \langle \mu \rangle = \frac{1}{n} \sum_i x_i, \quad \langle \sigma^2 \rangle = \frac{\sum_i (x_i - m)^2}{n-1} \quad \dots 2) \text{ 注2}$$

の式を使って、 μ と σ の値を、それぞれ、計測値から推測する。 x_i は、1つ1つの実測値 (i 番目の値という意味、標本^{注1})、 n は測定の回数 (標本数) を示す。 $\langle \mu \rangle$ および $\langle \sigma \rangle$ は、元の確率分布 (式1) 中の μ 、 σ とは厳密には同じではなく、あくまで推測された値である。もとの μ や σ の予測値 (期待値という) に過ぎない。 $\langle \rangle$ は期待値で、統計学的な処理によって得られた推定値であることを示す。エクセルの=AVERAGE (数値 1、数値 2、...) は $\langle \mu \rangle$ を、=STDEV (数値 1、数値 2、...) は $\langle \sigma \rangle$ を求めるときに使用する関数である。

S 測定誤差

さて、測定値には何らかのまちがいや誤差が必ず含まれる。例えば、図の測定値A (×印、破線の確率分布) の場合、平均の推定値 $\langle \mu \rangle$ が元の平均 (μ) とほぼ同じであるが、測定装置の何かの原因で、データのランダムな読み取り間違いが起こっているケースである。より大きく広がった分布となっている。このような測定を**精度 (precision)**の悪いという。しかし、この例では、平均値 $\langle \mu \rangle$ に関しては、もとの μ をほぼ正しく反映している。このような測定を、**正確さ (accuracy)**の面では問題のない測定であるという。測定の精度は、

$$s^2 = \frac{\sigma^2}{n} = \frac{\sum_i (x_i - \mu)^2}{n(n-1)} \quad \dots 3)$$

で計算される s を使って表現する。 s は**標準誤差 (S.E.、standard error)**とよばれるパラメータで、もとの平均値 μ を推定するときの精度を示すものである。複数回の測定によって、どの程度の精度で平均値が推定できるかを示す。上の式3に示すようにで、測定の回数 n を多くすればするほど (測定を繰り返す) と分母が大きくなるので、 s は小さくなる。これは平均値を求める精度が高くなったことを意味する。無限に測定を繰り返すと、 s は限りなくゼロに近づく。前の式2で計算される $\langle \sigma \rangle$ とはまったく意味の異なるパラメータである点に注意してほしい。一般に、実験で得られた結果を

たとえば、 3.4 ± 0.2 ($n=14$) と

$\mu \pm s$ (n =測定数)、または、 $\mu \pm \sigma$ (n =測定数)

のように、平均値±標準偏差 (σ) または標準誤差 (s) の形で表記する習慣である。ここで、式2による標準偏差 $\langle \sigma \rangle$ 、あるいは、式3による標準誤差 s のどちらを用いるのが適切であるかを正しく判断し、それを明記しなければならない。生命科学の分野では、観察対象のばらつきが大きく、それを正確に記述することが多い。標準誤差を議論する場面はむしろ例外的であろう。データは 3.4 ± 0.2 (SD) ($n=14$)と記述のもっとも一般的である。

例で考えてみよう。多数のタマネギ根端細胞について、そのサイズを顕微鏡観察で測定して表記する場合には、どちらの表記方法がふさわしいだろうか、また、あるタンパク質溶液の光吸収の量（吸光度）を繰り返して測定した場合どちらを選択すべきか、それぞれ、上の議論から判断できるであろう。もともと、ばらつきのある試料を多数あたって、その分布を推定して表現するのが $\langle\sigma\rangle$ である。本来は1つの値であるべきもので測定の都合上ばらつきが生じる場合、あるいは、計算した平均値の精度について記述しなければならないとき、 s を使わなければならない。また、このときに表記する桁数は、桁数が多く細かければ良いものではない。以下に述べる精度や誤差を考慮しての記述となる。

§ 系統誤差

上図の測定値B（○印、実線の確率分布）は、別の種類の測定誤差の見られるケースである。この例のように、ランダムなばらつきではなく、もとの平均値から一様に、片方向へとずれるような誤差を系統誤差（systematic error）という。この測定例Bでは測定の精度は高いが、本来の分布の右側に決まって値を読み間違える傾向がある。その結果、平均値を推定する場合、正確さを欠くことになる。このような誤差が発生する要因には、次の2つが考えられる。一つは測定装置の不安定性の問題、もう一つは、生体試料に独特の経時変化である。

前者は、装置の正確さを向上させる工夫を行うことで解決しなければならない。一般に絶対値が正確にわかっているもの（基準となるものさし、標準の分銅やおもり、基準濃度の溶液など）を使って、測定方法の校正（calibration）を行う。校正によって、元の値と測定値の間の補正を行った上で、実験に臨むことになる。測定装置が不安定で経時的な変化をどうしても除去できないこともあり、完全な解決の難しいことも多い。吸光光度計などの機器では、スイッチを入れて10～15分ほど待ってから測定を開始するが、これは装置を安定化させ、測定値の経時変化の影響をできるだけ小さくするためである。室温の変化や空調機のオン・オフが、同じように経時的な変化を起こす原因となることもある。

系統誤差のもう一つの要因は、生体試料そのものに由来することが多い。生きた生体試料を観察する場合、観察している間に時々刻々と生き物の状態が変わる可能性を考慮しなければならない。試料の移動・変形・疲労・老化・分解・変性・失活など活性状態の変化、熱発生・乾燥・吸水など観察条件に起因する人為的な変化などがある。ホルマリン固定などの薬剤処理や凍結・乾燥などの操作によって、その様な経時変化を抑えることもできるが、その処理そのものの影響が無視できないことも多い。このような問題点は、生体試料を使った実験では常に注意を払う必要がある。正確で再現性のよいデータを得るためには、実験の種類ごとの工夫が必要である。

§ 有効数字

どのような実験でも、上に述べたような試料そのもののばらつき（標準偏差 σ ）に、さらに測定誤

差が上乘せされる点に注意を払わなければならない。ある測定値を表記するときに、誤差の影響を受けない数字を有効数字 (effective digit) とよぶ。例えば、ある一つの細胞の長さを測定して、測定値の平均長が $250.616 \mu\text{m}$ 、標準誤差が $2.562 \mu\text{m}$ と式 2, 3 で求めた場合、平均値 \pm 標準誤差は、 $250.616 \pm 2.562 \mu\text{m}$ となるかもしれない。しかし、標準誤差を考えると 1 桁目の数字は測定を繰り返すたびに变化すると予想され、意味のない無効な数字となる。この場合、 $250 \pm 3 \mu\text{m}$ と表記するのが正しい。有効数字は 2 桁となる。しかし、ここでは生体試料の標準偏差についての議論が抜け落ちている。多数の細胞を観察し、その平均が 250.616 、標準偏差が 32.516 であったとする。この場合、測定の誤差の大小を細かく議論することはあまり意味がない。生体試料としての正規分布の広がりや正確に表記すること、どのような大きさの分布があるのかを記述する方が正しい。測定の誤差を考慮して、 $250 \pm 33 \mu\text{m}$ (標準偏差) と表記することが生物学的には正しい表記方法である。

注 1 正確には、測定の対象となるグループを母集団、その分布の μ 、 σ を、それぞれ、母平均 (population mean) と母標準偏差 (population standard deviation) とよぶ。実験データは、その母集団から、ランダムに選んでデータを得たもので、これを標本 (sample) という。標本をもとに推定した $\langle \mu \rangle$ と $\langle \sigma \rangle$ を、それぞれ標本平均値 (sample mean)、標本標準偏差 (sample standard deviation) とよぶ。

注 2

$$\begin{aligned} \sum_i (x_i - m)^2 &= \sum_i (x_i^2 - 2m \cdot x_i + m^2) = \\ &= \sum_i (x_i^2) - 2 \sum_i (m \cdot x_i) + \sum_i (m^2) = \sum_i (x_i^2) - 2m \sum_i (x_i) + n \cdot m^2 = \sum_i (x_i^2) - n \cdot m^2 \end{aligned}$$

と式の上では変形できるので、手計算で標準偏差を求めるときは、 $\langle \sigma^2 \rangle = \frac{\sum_i (x_i^2) - n \cdot m^2}{n-1}$ を使うと便利である。

演習課題

1. 次の標本平均値、標本標準偏差、標準誤差を求めなさい。

0.779, 0.633, 0.394, 0.994, 0.8, 0.194, 0.779

0.008, 0.936, 0.339, 0.794, 0.329

2. 次の測定値のばらつきは、標準偏差 (SD) で表現すべきか、標準誤差 (SE) で表現すべきか。

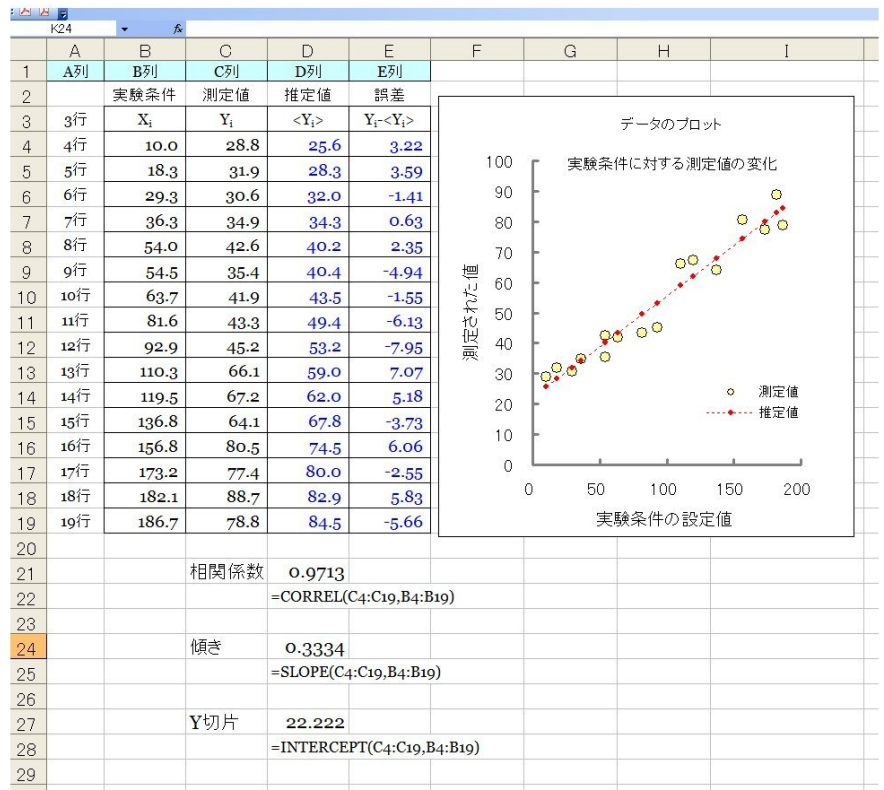
- a. マイクロピペット P100 (100 μ L) で、100 個マイクロチューブに分注したときのそれぞれのサンプル溶液の量
- b. 最後にマイクロピペット P100 (100 μ L) で吸い取った試料溶液の重量を 10 回精密天秤で測定した結果
- c. ある学年の男子学生全員の身長データ
- d. 自分の身長を 3 分おきで計測した 24 時間分のデータ
- e. 春日通りを通過する都バスに乗車している乗客の数の分布
- f. 10 匹のマウスに薬品 A を投与し続け、1 ヶ月後の体重と実験開始前の体重の比率
- g. 1g の標準質量の分銅を、20 回精密天秤で測定した結果
- h. カエル座骨神経の興奮を細胞外電極で 10 回刺激して記録したときの振幅データ
- i. 100 μ g の重さのタンパク質を 1 mL の水溶液に溶解した溶液の紫外線吸収量を 10 回計り直した結果。
- j. サイコロを 100 回振って出てくる目の数の分布
- k. ウニ卵を 100 個受精させて、それぞれで最初の卵割が起こるまでの時間

最小二乗法と直線回帰

実験では、条件をさまざまに変え、そのとき生体の反応はどのようになるか調べることで、さらに詳細な理解が得られることが多い。例えば、生物に投与する薬剤の量 (x) を少しずつ変えたときにどのような反応 (y) の差が現れるか、刺激の強さ (x) を変えると観察される神経の興奮の大きさ (y) はどのように変わるか、溶液に加える基質の量 (x) を変えると反応の速度 (y) はどのように変わるか・・・など、 y は x でどのように変わるかの予測をおこない、それを確かめる実験を行う。実習でもこのような実験例を多数学んだであろう。ここでは、そのような実験データで一般に使われる直線回帰の方法と最小二乗法について解説する。科学実験では非常に基本的な解析方法である。原理を正しく理解し、エクセルによる計算方法も習得しておくとう便利である。

§ 直線回帰の実例

まず、実例を紹介して、その後で原理を解説する。右図は、実験条件を変えたとき (X_i : 設定した温度や濃度など)、結果 (Y_i : 測定値、○) がどのように変化するかを示したグラフである。 X_i の i は、何番目の測定条件であるか、 Y_i の i は、何番目の測定結果であるかを示す。この実験では、ばらつきはあるものの、その分布から、 X_i と Y_i の関係は直線関係にあると考えられるデータである。精度の善し悪しはあるだろうが、 $Y = aX + b$ の形で表現できれば、次に新しく X_i を設定して実験せずとも、 Y_i の



値を予測することもできる。この直線を求める方法が、**最小二乗法** (least square method) である。理論直線を求める作業を**直線回帰** (linear regression)、その理論直線の正確さを示す指標が**相関係数** (correlation coefficient) である。 X_i と Y_i の関係をこのような方法で解析する作業を、「**相関** (correlation) を調べる」、上のような X_i と Y_i の関係を示すグラフを**散布図** (scatter diagram) という。

§ エクセルを使った計算

エクセルの中の統計専用の関数には、平均値や標準偏差など、公式を使わなくても簡単に計算する関数が多数含まれている。同じように最小二乗法も統計用の関数を使うと便利である。 $Y = aX + b$ と直線をおいたときの、 a を傾き (slope)、 b を Y 切片 (y-Intercept) とよび、それぞれ、

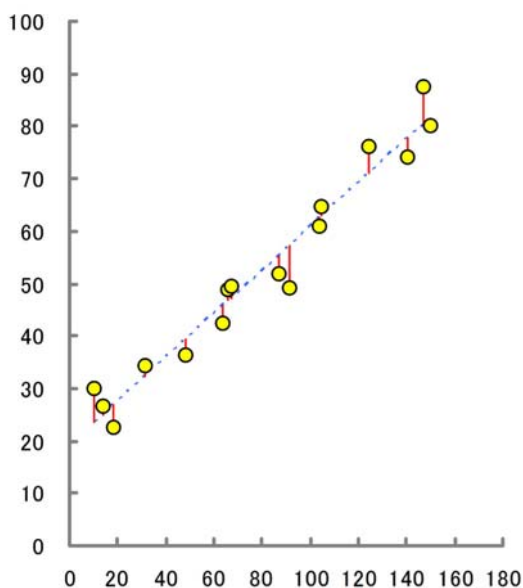
$$m_X = \frac{\sum X_i}{n}, \quad m_Y = \frac{\sum Y_i}{n}, \quad a = \frac{\sum X_i \cdot Y_i - n \cdot m_X \cdot m_Y}{\sum X_i^2 - n \cdot m_X^2}, \quad b = m_Y - a \cdot m_X \quad \dots 4)$$

で計算する。 m_X と m_Y は、 X_i と Y_i 、それぞれの平均値である。相関係数は、

$$R = \frac{\sum X_i Y_i - n \cdot m_X \cdot m_Y}{\sqrt{\sum X_i^2 - n \cdot m_X^2} \cdot \sqrt{\sum Y_i^2 - n \cdot m_Y^2}} \quad \dots 5)$$

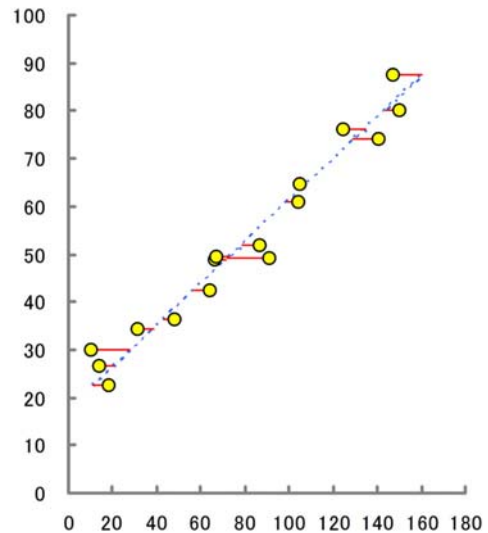
を使って計算する。エクセル上では、前図の中に示すように、傾きは「=SLOPE(C4:C19,B4:B19)」で、Y 切片は「=INTERCEPT(C4:C19,B4:B19)」で、相関係数は「=CORREL(C4:C19,B4:B19)」で計算する。このエクセルの () 内の列・行の表記方法は、エクセルの「ヘルプ機能」を使って確認した上で、使用するよう注意してほしい (C4 は C 列 4 行目のマス目 (エクセルでは「セル」とよぶ) を指す)。

ここで計算した a と b を使い、 $\langle Y_i \rangle = aX_i + b$ の推定値を計算する (下図内の破線)。 $\langle Y_i \rangle$ は、 Y_i の推定値 (期待値) の意味である。推定値 $\langle Y_i \rangle$ と、実際の測定値 Y_i の値との差、 $Y_i - \langle Y_i \rangle$ が、推計の誤差となる。図では、この $Y_i - \langle Y_i \rangle$ を縦方向の短い実線で示している。



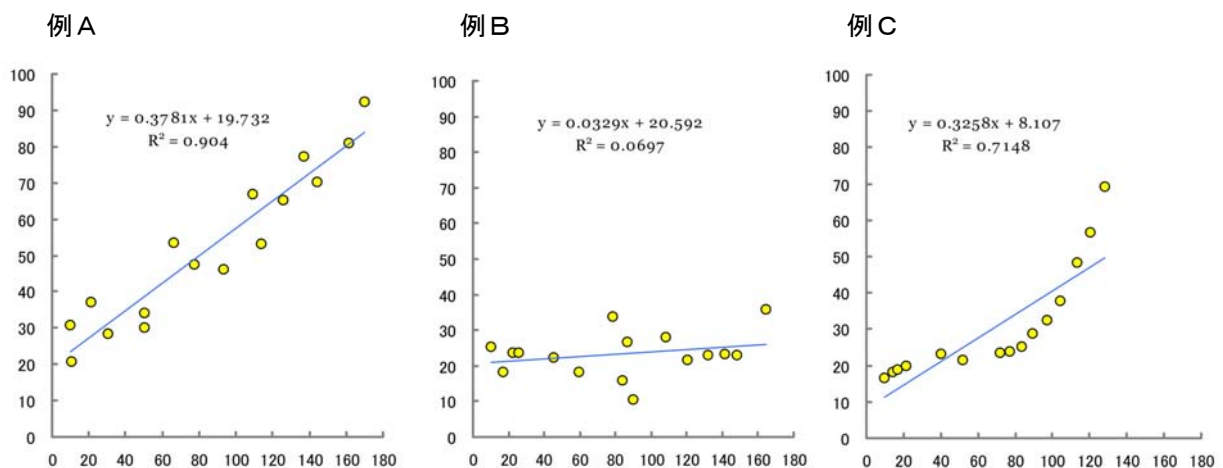
§ 最小二乗法の原理

最小二乗法とは、実測値と推計値の間の誤差が一番小さくなるようにしたもので、これは一番データに近い直線回帰であることが数学的にも証明されている。誤差を最小にするのには、それぞれの測定点で $(Y_i - \langle Y_i \rangle)^2$ を計算し、その総和、 $s = \sum (Y_i - \langle Y_i \rangle)^2$ がもっとも小さくなるように a と b を決める。 s は、上の図では、縦方向の $Y_i - \langle Y_i \rangle$ の線分長の二乗和に相当する。 $\partial s / \partial a = 0$ 、および $\partial s / \partial b = 0$ となる条件 (a や b を変えたときに s が極小値となる条件) を探すと、 a や b の答えが得られる。なぜ、上のような式 (4) を使うと a や b が計算できるのかは、統計学の専門書を参照してほしい。この式の導出は、さほど難しい作業ではないだろう。また、右側の図のように、横方向に $(X_i - \langle X_i \rangle)^2$ を計算し、 $s = \sum (X_i - \langle X_i \rangle)^2$ が最小となるように $X = aY + b$ の直線に回帰する計算を行うと、その結果はもちろん大きく異なってくる。これは、先に説明した計算例とは、何が、本質的に違うか考えてみよう。



§ 直線回帰の精度

最小二乗法の計算は、実測値と推計値の間の誤差が一番小さくなるようにしたもので、一番データに近い直線となり、相関係数 R は、 X_i から Y_i を推定するときの「正確さ」を示すパラメータとなる。例を3つ紹介する。



例Aは、一般的な解析例で、ばらつきは大きいですが、 X_i と Y_i は直線的な関係（片方が変われば、それにつれて直線的に変化する）があることを示している。 R^2 が1に近いとき、「強い相関がある」、0に近いときに「相関が低い」という判断を行う。ところが、解釈を行う上で、注意しなければならない点もある。例Bは、例Aとばらつきの程度はさほど変わらないが、 R の値は大きく異なる。これは、 X_i の値から Y_i を推計することが大変難しいことを意味する。傾き a が小さな場合、測定の誤差とは関係なく、相関係数は小さくなるケースである。例Cも、相関係数 R^2 は0.7と例Aよりは低いですが、 X_i と Y_i の間には明瞭な相関が見られるケースである。しかし、直線関係ではない。この場合には、 $Y = aX + b$ の直線への回帰ができないので、 $Y = aX^2 + bX + c$ や $Y = aX^3 + bX^2 + cX + d$ 、あるいは、もっと複雑な理論式に回帰させる必要があるだろう。このような計算は、前ページの計算式のように公式を使って a や b を求めることはできない（これを、「解析的に求められない」、「解析解がない」という）。コンピュータを使って、計算を試行錯誤して、 $s = \sum (Y_i - \langle Y_i \rangle)^2$ がもっとも小さくなる条件を見つける（これを「数値解を求める」という）。'R'や'SALS'などの専用の統計プログラム、あるいは、次節のエクセルのソルバー機能を使うことになる。このように、式5で計算する相関係数 R の意味は、直線回帰であることが前提である点は十分注意しなければならない。

実験データを解釈する上で、相関係数 R は数学的な意味でしかない。必ずしも、因果関係を証明していない点にも十分に注意しなければならない。例えば、ある酵素反応の時間経過を調べているときに、0分、1分、2分と分解される量を調べると、明らかな相関が現れ、 $R^2 \sim 1$ となるケースが多いだろう。しかし、この相関関係から、「基質が酵素によって分解された」と結論することはできない。溶液の中に入っている基質が時間とともに自然分解する性質がある場合にも、同じような相関が観察されるからである。特に2つの変数の間に微妙な相関しか見られない場合にはじゅうぶんな注意を払う必要がある。偶然に、2つが同じ変動をすることはしないのか、慎重な判断や他の方法での検証が必要となる。

エクセルのソルバー機能を使った最小二乗法

最小二乗法を用いると、最適とされる近似式、たとえば、 $Y = aX + b$ の式に結果を当てはめるときの a や b を容易に求めることができる。同じ方法を、もっと複雑な関数、たとえば、もっと複雑な関数、 $Y = aX^2 + bX + C$ や $Y = a \cdot \exp(b \cdot X)^2 + cX + d$ などへも原理的には応用できる。問題は、前節で説明したような簡単な数式で $a \sim d$ などの定数を表現できない点である。こういった関数は、簡単な一次関数の足し合わせの関数（これを線形の関数とよぶ）ではないので**非線形最小二乗法**という。

非線形関数の最小二乗法は、かつては専用の解析ソフトやプログラムを使って解決しなければならなかった、現在のマイクロソフト社のエクセルには、**ソルバー（解法）**というメニューが用意されていて、それをうまく使用すると容易に非線形の最小二乗法が可能となる。その計算例を紹介する。

§ソルバー使用の実例

実際の操作は、エクセルのワークシートを使って解説する。まず、下のエクセルファイルをダウンロードして、コンピュータで開く。

www.bio.chuo-u.ac.jp/nano/books/SolverSample.xls

このエクセルファイルの中にも解説を記してあるが、下がその概要である。

まず、はじめに実験の結果をどのような関数で近似したいかを決める必要があり、この関数の表記方法は、エクセルの指定した方法にしたがう。たとえば、

$$Y = a \sin bX + cX^2 + d \quad \dots 6)$$

などのかなり複雑なものでも

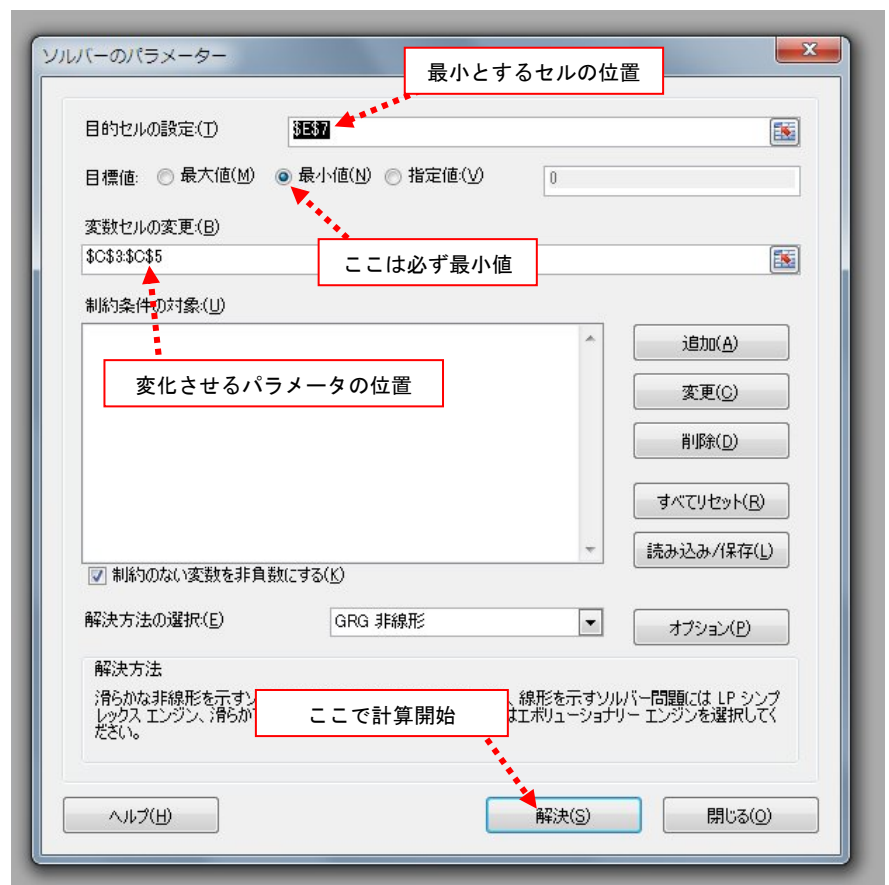
$$= \$C\$3 * \text{SIN}(\$C\$4 * B10) + \$C\$5 * B10^2 + \$C\$6 \quad \dots 7)$$

と書くだけでよい。この式の表記上の約束は、B10 が B 列 10 行目のセルに書かれたデータ X である。 $\$C\3 、 $\$C\4 、 $\$C\5 、 $\$C\6 は、上の式中の $a \sim d$ に対応する定数で、C 列の 3～5 行目のセルに書かれた値をここでは使うという約束事である。この場所は、他の任意の場所に変えて指定しても良い。

「ソルバー」の機能をスタートさせると、この定数値を変えながら、探し出すべきベストな解を自動的に見つけていく。

ここで記述した関数を使って、 X_i のデータ (B 列) から Y_i を計算し、これを予測値とする ($\langle Y_i \rangle$ とここでは表記)。 Y_i は別個に実験で求めた値なので、その誤差 ($Y_i - \langle Y_i \rangle$) の二乗を計算し、その合計をエクセルの表の中で、=SUM という関数で計算する。上のエクセルファイル例では、その二乗和の値は、E 列 7 行目に書き込むように指定されている。これで準備完了である。あとはソルバー機能を開始するだけでよい。その手順 (一次関数の例) を下に記す。

- ① まず、最初に関数を確認する。エクセルファイルの「一次関数」のワークシートでは、一番簡単な数式 ($Y = aX + b$) が、「B10*\$C\$3+\$C\$4」、2 つ目は「B11*\$C\$3+\$C\$4」、3 つ目は「B12*\$C\$3+\$C\$4」の形で表記されている。これは C 列 3~4 のデータ (a と b) をいつも使って、B10、B11、B12・・・と順番に X_i を使って計算することになる。この表記方法はエクセル独特の約束事で、これを覚えた上で正確に記述する必要がある。たとえば、A10 と \$A\$10 と書いた場合、これは、エクセルシート上でコピー・ペーストしたとき、それぞれ、セル位置の移動で変わる部分 (A10) と、固定される部分 \$A\$10 の指定方法の違いである。
- ② E 列 7 行目に=SUM で、誤差の二乗和合計が正しく計算されていることを確認する。
- ③ 次に「データ」をクリックして、ソルバーの表示 (右) を出す(Excel2010)。エクセルのバージョンによっては「データ」→「分析」でソルバーを選ぶ。ただし、あらかじめ「Excel のオプション」→「アドイン」→「設定」で ソルバーアドインを使うという設定にしておく必要がある。
- ④ 「ソルバー」をクリックすると右の画面が表示され、ここでパラメータを確認して、「解決」のクリックで計算を実行する。ここでのパラメータの設定方法の詳細は、ダウンロードしたファイルの中で、「一次関数」のワークシートに書かれた手順書を参照にする。



§ソルバーの応用

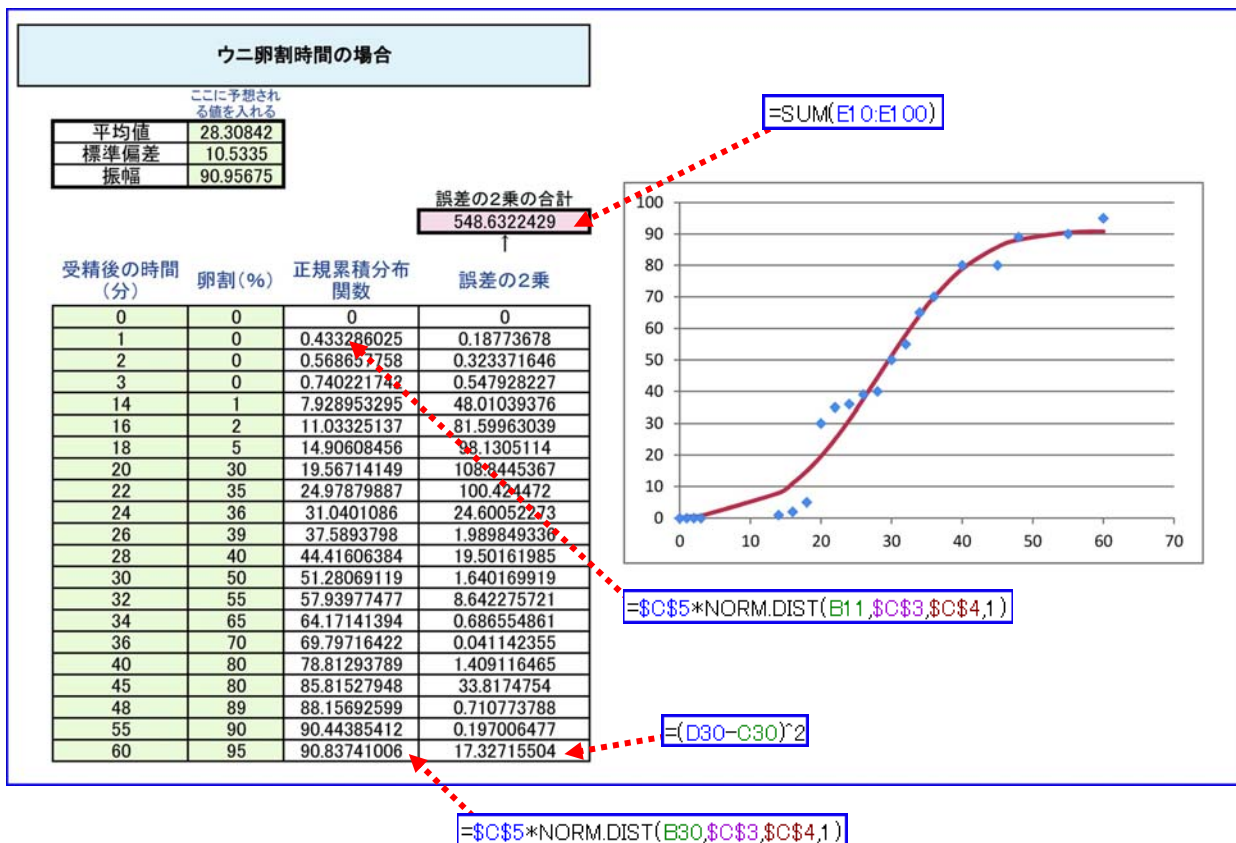
エクセルのソルバーは、かなり複雑な関数であっても最小二乗法を実施できる。実験データを解析する上で大変便利な機能である。上でダウンロードしたファイル内には、ウニ受精卵の卵割のタイミングを調べた実験結果を正規累積分布関数に近似して、平均卵割時を推定する例題も用意されている(下のワークシート)。

正規累積分布関数は、一般に

$$erf(X) = a \int_0^X \exp\left(-\frac{(T-b)^2}{2c^2}\right) dT \quad = \$C\$5 * NORM.DIST(B10, \$C\$3, \$C\$4, 1) \quad \dots 8)$$

と表現される複雑な関数で、正規分布(「①測定と誤差について」参照)を積算したS字型カーブである。エクセルの上では、上の右側のような表記で書き、ソルバーを使った解法が可能となる。このように、ソルバーの応用範囲は無限と行って良いほど広い。今回の実習の機会に習得することをおすすめしたい。サンプルのエクセルファイルの中には、指数関数への回帰の例も含めた。

エクセルシートの中のウニ卵割時間を求める例題(赤破線矢印はセル内の表記例)



練習課題(1) 下のデータは、ウニ胚の第一卵割を起こした比率 (%) をある時間間隔で調べた結果である。このデータをもとに、上のソルバー機能を使って、ウニ胚の平均卵割時 (受精後何分か) を求めよ。

受精後の時間 (分)	卵割が完了した 卵の比率 (%)
0	0.0
5	0.0
10	5.1
25	10.0
30	18.3
40	27.7
45	44.7
50	46.0
55	69.3
60	76.7
70	84.0
75	90.0
80	95.0

練習課題(2) 上では卵割を起こした比率 (%) を一定時間おきに調べて、それを正規累積分布関数に近似する方法を紹介した。しかし、実際は、ある卵に注目して、それが受精後何分で第一卵割を起こしたかという時間のデータを蓄積する方がデータを得やすいかも知れない。その場合、データは

卵の番号	第一卵割	第二卵割
1	45	61 (分)
2	40	58
3	42	50
4	46	65
⋮	⋮	⋮
⋮	⋮	⋮
⋮	⋮	⋮
⋮	⋮	⋮

となると考えられる。このようなデータを直接使って、ソルバー機能を活用し、平均の第一卵割時間や第二卵割時間を求めるにはどのようなデータ処理するのがよいだろうか (もちろん、このようなデータが十分な数あれば、卵割%に表現し直し、上のワークシートで同じように解決できるはずであるが) ? どのような関数をワークシートの中で定義して、ソルバーに解かせるのが良いか?