有効積算温度法則パラメータの新しい推定法

帝京大学医学部微生物学教室　池(いけ)本(もと)孝(たか)哉(や)
聖マリアンナ医科大学免疫・病害動物学教室　高(たか)井(い)憲(けん)治(じ)

はじめに

有効積算温度法則は害虫をはじめとした変温動物などについて，それらの環境温度（T°C）と発育期間（D日）の間に，

$$D(T-t)=k \qquad (1)$$

なる関係が成り立つことをいうものである。ここでtは発育ゼロ点，kは有効積算温度定数と呼ばれるパラメータである。(1)式を変形して直角双曲線の形にした

$$D=\frac{k}{T-t} \qquad (2)$$

として視覚的に理解することもできる。しかしこのままでは観察データを当てはめるのは困難であり，これまで一般的には実験観察値から(1)式を変形した

$$\frac{1}{D}=-\frac{t}{k}+\frac{1}{k}T \qquad (3)$$

なる直線式に回帰分析を適用してこれらのパラメータ値が推定されてきたのであるが，実験観察値が完全に直線とならず多少ともバラツキがある場合には，推定値は偏りの大きいものとなり信頼性の低いものであることが分かった。読者の中にも，低温部のデータ点が推定した双曲線から大きく外れ，不快な思いをされた経験をおもちの方がおられることと思われる。

I　従来法のどこが問題か

1　低温部のデータ点がかなり無視されて直線化される

従来の方法によって求めた回帰直線のまわり（つまり上下）に標準誤差を描き入れてみよう。数式は

$$s(1/D)\equiv s_e=\sqrt{\frac{\Sigma(Y_i-y_i)^2}{n-2}}$$

であり（Yiはデータ点，yiはこれに対応する直線上の推定点でデータ点より直線までY軸方向に移動したもの，nはデータ数とする），上下線は

$$\frac{1}{D}=-\frac{t}{k}+\frac{1}{k}T\pm s_e$$

となる。つまりデータ点の約68%がこれら上下線の中にあるはずである。(4)式をもとの双曲線である(3)式の形に戻すと

$$D=\frac{k}{T-(t\mp k\cdot s_e)}$$

と，2本の双曲線となる。図-1に示したように，推定した双曲線まわりの標準誤差は温度が低くなるにつれて急に大きくなっている（分散の不均一性）。さらに定規を縦に当てていただきたい。ある温度における標準誤差は推定双曲線の上側で大きく，下側で小さいのである（分散の非対称性）。（なお実際に直線まわりの標準誤差を計算するには図-6の数式一覧にある数式を使って求める）

これらの結果をデータ点から双曲線を推定するという作業の内容としてとらえると，高温部分のデータ点ほど重要視され，低温部分のデータ点ほど軽んじられるか，ほとんど無視されていることになる。さらに同じ温度であれば，推定双曲線の下側のデータ点は上側のそれより重要視されたのである。実際の場面でいうと，せっかく苦労して測定した低温部の発育期間のデータが，ほとんど無視された形で発育ゼロ点tと有効積算温度定数k

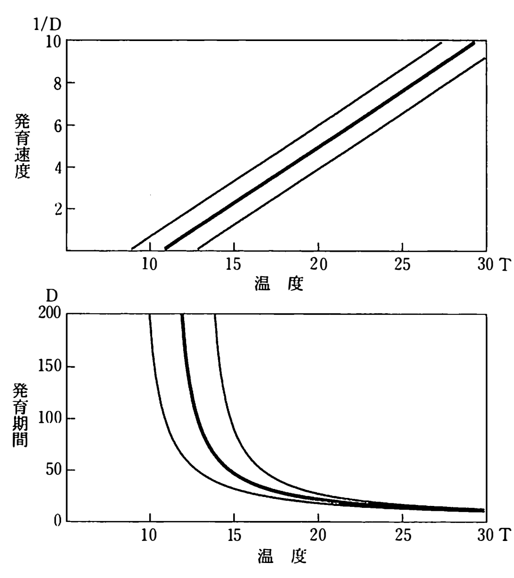


図-1　$1/D=-(t/k)+(1/k)T$ 式にy on x回帰直線を当てはめたときの直線まわりの標準誤差

A New Method for Estimating the Parameters of the Law of Total Effective Temperature.　By Takaya IKEMOTO and Kenji TAKAI

（キーワード：有効積算温度，発育ゼロ点，有効積算温度定数，直線化，不偏長軸，パラメータ，回帰直線）

のパラメータが推定されており，しかも推定値には偏りがあることとなる。

2 温度データに誤差があると推定直線は寝てしまう

温度設定した恒温室で昆虫の発育日数を調べた場合，各個体の発育期間にはバラツキが見られる。つまり平均値とその標準誤差という形でデータ点が得られ，発育期間 D は誤差をもつことになる。

では温度データ T には誤差がないのであろうか。最も分かりやすいのは野外の自然条件下で測定したデータであり，必ず誤差が存在する。恒温室や飼育容器内の温度も正確に設定されても，外部気温の変動に影響されて，少しは変動しているのである。蛇足であるが，設定した上で，さらに測定されることを期待したい。そうすれば平均温度とその標準誤差が測定されるはずである。

ところで，普通の直線回帰分析ではX軸（T軸）変数（温度）には誤差があってはならないことになっている。そうでないと直線の傾きも切片の値も求めてはいけないのである。もしこの誤差を無視して直線回帰するとどうなるだろうか。証明は統計学書（たとえばスネデッカー・コクラン，1967）にゆずるとして結果だけをいえば t は小さく，k は大きく推定されてしまうのである。

II いろいろな直線化（line-fitting）法

1 回帰直線法

直線的な傾向が見られるデータポイントに直線を当てはめる方法といえば，直線回帰法しかないかといえば決してそのようなことはない。むしろこれは特殊な場合に使ってよい方法の一つである。すでに述べたように，これを使うには条件があって，X軸変数にまったく誤差がない場合である。たとえば，年次的に害虫の数が直線的に増加していることを示したいとき，X軸変数にはカレンダー時間をとることになるので，この条件は満たされている。

2 長軸（直交回帰直線）法

もし，A，B両社製の濃度測定器があって，いろいろな濃度をもつ試料をそれぞれの測定器で測ってグラフ用紙にプロットしたとき，多少のバラツキがあっても直線性が見られ，直線の傾きが1（45度）であれば両社製の測定器に性能の差はないと見ることができる。このような場合に適用できるものに，長軸（直交回帰直線ともいう）がある。回帰直線が各データ点から推定直線までY軸方向に引いた線分をその1辺とする正方形の面積の総和を最小とする直線であるのに対し，長軸は各データ点から推定直線まで引いた垂線を辺とする正方形の面積の総和を最小とする直線である。

回帰直線しか念頭にない人にとっては驚くべきことに，このようなデータに長軸を適用して直線の傾きが1となった場合，同じデータに普通の直線回帰を適用すると傾きは1より小さく推定される。

3 不偏長軸法

昆虫各個体の翅長と体重を測定すると，翅長の大きい個体は体重が重いという直線傾向があり，この関係を直線式で示したいときは，不偏長軸という直線を用いることが推奨される（Kermack and Haldane, 1950）。これは各データ点から推定直線まで，Y軸方向とX軸方向に線を引き，これらを2辺とする長方形の面積の総和を最小とする直線である。長軸法では変数の単位の取り方（たとえばgまたはmgでとったどちらの数値で計算するかによって）推定直線の傾きが変わってしまう欠点があるのに対し，不偏長軸法は変数を標準化してしまうので，そのような欠点がない。

数理統計学的にはその他にもいろいろな数学的モデル式が開発されており，たとえば，λ 直線という長軸や不偏長軸を包含した直線モデルが提示されている。しかし2つの変数の分散比があらかじめわかっていないと推定には使えないなど，実用面で問題がある。現在実際に使えそうなこの3種類の直線を図-2とともにもう少し整理しておこう。

回帰直線（linear regression of y on x）：Y軸方向の残差平方和を最小にする直線で，2変数のうちY軸変数だけに誤差が認められる。どちらをY軸変数とするかで2本の推定直線が得られる。（著者らはこれをY軸残差直線，Y軸残差法と呼んではどうかと考えている）

長軸（major axis, orthogonal regression line）：推定線と垂直方向の残差平方和を最小にする直線で，両変数に誤差があってもよい。2変数確率分布を示す楕円の長軸として与えられる。変数をどちらにとっても同じ直線が求まる。（同じく，垂線残差直線，垂線残差法）

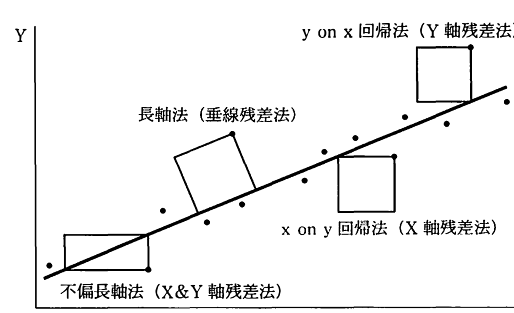


図-2 残差を最小にする四つのline-fitting法
各方法において4角形の面積総和を最小にする

不偏長軸（reduced major axis）：長軸とほぼ同じであるが，両軸変数が標準化される。この直線の傾きの係数は2本の回帰直線（y on x 直線と x on y 直線）のそれぞれの傾きの係数の幾何平均としても求められる（Cheng and Van Ness, 1999）。図-6の数式を参照されたい。（同じく，X & Y 軸残差直線，X & Y 軸残差法）

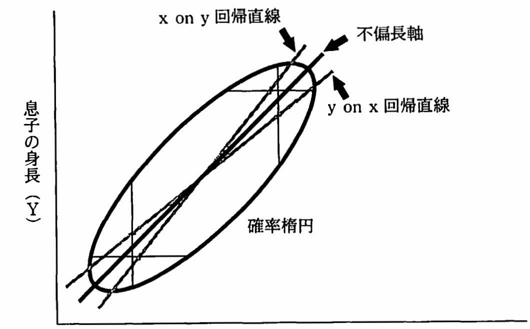
III　不偏長軸法（X & Y 軸残差法）の効用と弱点

これまでの説明では，不偏長軸法が最も一般性があり，変数のもつ誤差など数学的な性質をほとんど気にすることなく使え，極端には回帰直線を使うことをやめてもよさそうである。

ところが不偏長軸法には数理統計学上の弱点がある。その最も重大なものは，不偏長軸は2変数の変数内誤差モデルの最尤解（必要条件）として得られるが，十分条件は満たしていないことと，無限母集団における推定直線の傾きが，真の傾きに理論上一致しないことである（Solari, 1969）。このことは厳格さを尊ぶ数理統計学者に嫌われ，擁護派の学者たちとの論争がここ何十年と続いている。擁護派の Ricker（1975, 1984）によれば，一致性の基準はあくまでも無限母集団での性質であり，一致性があるからといってそれが有限な標本集団でのよい推定量を保証することにはつながらないから，その逆もまた真なりであり，ことさら問題視すべきでないと主張している。

ちなみに普通の回帰直線は，実際の場ではそのようなことはほとんどないにもかかわらず，X 軸変数に誤差がまったくないと仮定しておくと，このような理論的問題点はすべて解消される。したがって現在，大多数の一般向け数理統計学書には回帰直線のことが詳しく解説されており，不偏長軸法はほとんど取り上げられていないのが現状である。しかしながら回帰直線にも弱点があり，推定に使える X 軸変数の範囲はデータ点が存在する内挿的な部分に限られ，有効積算温度法則の発育ゼロ点のように外挿的な部分の推定には大きな誤差が出る確率が高い。これに対して不偏長軸は内挿，外挿ともに妥当な推定値が得られるのである（Ricker 1975, 1984）。

そもそも回帰（regress）という概念はイギリスの遺伝学者 F. ガルトン（1822～1911）にはじまる。彼は1,078組の親子の身長を分析し，父親の身長を階級分けして X 軸に，その息子たちの平均身長を Y 軸にとったとき，これらの関係を示す直線の傾きは1より小さいことを見いだした。身長の高い親の息子はやはり高いが，父親ほど高くなく，身長の低い親の息子はやはり低いが，父親ほど低くない。つまり，平均身長から外れた父親の息子の身長は，世代を重ねて平均へと回帰（regress＝go back）する，退行の法則を提唱したのであった。



図－3　確率楕円に対する回帰直線と不偏長軸の違い

ガルトンが行った数値分析の手順に，ガウス（1777～1855）の最小2乗法を組み合わせると，現在で言う回帰分析法そのものになる。すなわち1,078組の X-Y 関係を回帰直線分析すれば，その直線の傾きは間違いなく1より小さくなり，ガルトンが示した数値とほぼ一致するはずである。同じデータを不偏長軸法で分析するとどうなるであろうか。直線の傾きがほぼ1になると推測できる。回帰直線と不偏長軸ではこれほど意味が違うのである。また，現在のイギリス人男性の身長は，決して平均的な人ばかりでなく，昔と同じようにばらついており，将来においても平均身長へ回帰すると考える人はまずいないであろう（図-3）。

さて，本論の有効積算温度に戻ろう。ここで回帰直線をやめて，不偏長軸を採用すると決めても，I 章で議論した（3）式を使って推定することにまつわる問題点は未解決のままである。

IV　新しい直線式と不偏長軸法との組み合わせ

1　新式 $(DT)=k+tD$ による問題点の解消

これまで普通に使われてきた（3）式に代わるものとして Ikemoto and Takai（2000）は

$$(DT)=k+tD \qquad (4)$$

式を提案した。Y 軸変数は発育期間と温度を乗じたものであり，X 軸変数は発育期間である。（3）式と同様に，この（4）式も（1）式の代数的な変形に過ぎないが，パラメータ k と t がそのまま直線の切片と傾きの係数となっている。（3）式について考えた直線まわりの標準誤差は

$$(DT)=k+tD\pm s'_e$$

であり，これを双曲線の形にすると

$$D=\frac{k \pm s_e'}{T-t}$$

となる（図-4）。図-2と比べていただきたい。分散の不均一性は大きく改善され，分散の非対称性は完全に解消されていることがわかっていただけると思う。つまり低温部のデータポイントも高温部のそれもほぼ同様に重要視されて二つのパラメータが推定される。

直線の当てはめに際しては，X軸変数が D であり，明らかに誤差をもつから，不偏長軸法を採用するのは言うまでもない。

なお，おことわりしておくが，(3)式に代えて(4)式を提案したのはline-fittingの場面で用いるためであって，(3)式の関数式としての重要性が失われたわけではない。

2　適温領域の決定に威力を発揮

有効積算温度法則が成り立つのはいわゆる適温領域内においてであり，それより高温になるとかえって発育速度が低下し，直線関係が成立しなくなる。低温域ではまた別の応答系が働くと見えて，ここでも直線性が崩れる。(3)式の直線が成立するのは適温範囲内のみであり，その両端の外まで考えると，一般にはS字カーブのようになることはよく知られている。そもそもカーブを曲がってしまった適温範囲外のデータ点はオミットしてline-fittingしなければならない。

実際上バラツキのあるデータから，適温範囲をどのように決定したらよいだろうか。筆者らは(3)式で描いたグラフよりも(4)式での方が明確にこの範囲を決められることを経験的に見いだした。

図-5を見ると(A)の(3)式で示した方で，低温領域の端の1点はやや直線からのズレが大きいと分かる。高温領域については端の2点が大きいズレを示している。ところが(B)の(4)式で示した方では，温度の順にデータ点を結んでみると，右上の低温領域では3点が明らかに直線から大きくズレていることが分かり，左下の高温領域の3点もやはり直線とは異なる傾向をもった点群であることが分かる。つまり，これら低温，高温領域の端の点を除いた範囲がこの昆虫の適温領域であるということになる。

3　パラメータおよびその誤差推定が簡単

従来の(3)式では，切片の値 α，傾きの係数 β から必要なパラメータ k，t を得るのに，もう2段の計算，

$$k=\frac{1}{\beta},\quad t=\frac{\alpha}{\beta}$$

が必要だが，新しい(4)式 $(DT)=k+tD$ では，α，β が即 k, t となり明解である。さらにこれらパラメータの

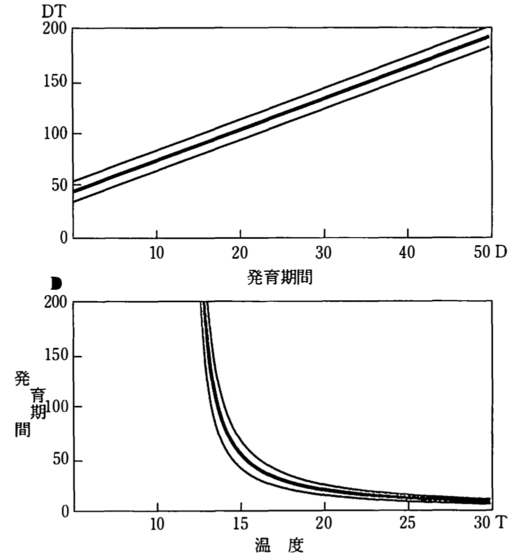


図-4　$(DT)=k+tD$ 式に不偏長軸を当てはめたときの直線まわりの標準誤差

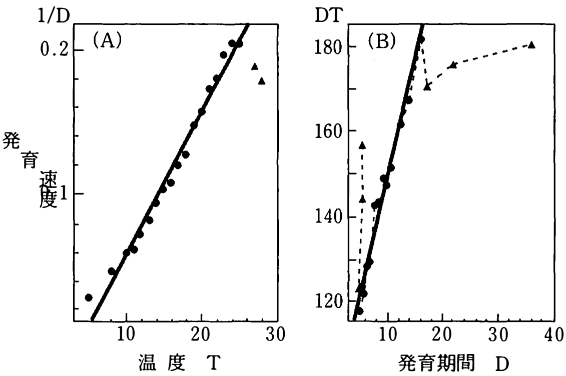


図-5　エンドウアリマキ *Arcyrthosiphon pisum* 幼虫の温度依存発育における適温範囲検出に関する2法の比較（データ：LAMB 1992）

誤差推定に至っては，前者の場合，誤差伝播式を用いて導かれた複雑な計算を要する（図-6の数式一覧参照）。後者の場合には，α，β の推定誤差が即 k，t の推定誤差となり市販の数表計算ソフトでそのまま計算できる。

V　数　値　例

読者の便宜のために簡単な数値例を表-1に示しておいた。図-6の数式一覧を参考にして実際に計算していただきたい。この数値例により実際に描いたのが図-1と図-4であり，数値例ではバラツキが小さいので，標準誤差線はそれぞれ標準誤差を3倍にして描いてある。従来の方法で推定すると発育ゼロ点 t がこれより約1°C低く，有効積算温度定数 k は約10大きくなる。

Line-fitting 数式一覧

直線式	$Y=\alpha+\beta X$	
定義式	$S_{xx}=\Sigma(X_i-\bar{X})^2$　$S_{xy}=\Sigma(X_i-\bar{X})(Y_i-\bar{Y})$　$S_{yy}=\Sigma(Y_i-\bar{Y})^2$ $r=\frac{S_{xy}}{\sqrt{S_{xx}S_{yy}}}$ (相関係数)	
分析法	y on x 回帰分析	不偏長軸分析
関係式	$\frac{1}{D}=-\frac{t}{k}+\frac{1}{k}T$	$(DT)=k+tD$
傾き 標準誤差	$\beta=\frac{S_{xy}}{S_{xx}}$, $s(\beta)=s_e\frac{1}{\sqrt{S_{xx}}}$	$\beta=\sqrt{\frac{S_{yy}}{S_{xx}}}$, $s(\beta)=\beta\sqrt{\frac{1-r^2}{n-2}}$
切片 標準誤差	$\alpha=\bar{Y}-\beta\bar{X}$, $s(\alpha)=s_e\sqrt{\frac{1}{n}+\frac{\bar{X}^2}{S_{xx}}}$	$\alpha=\bar{Y}-\beta\bar{X}$, $s(\alpha)=\sqrt{\frac{S_{yy}}{n}\cdot\frac{1-r}{n-2}\left(2+\bar{X}^2\frac{(1+r)n}{S_{xx}}\right)}$
直　線 まわりの 標準誤差	$s_e=\sqrt{\frac{S_{yy}-\beta S_{xy}}{n-2}}$	$s'_e=\sqrt{\frac{2}{n-2}(S_{yy}-\beta S_{xy})}$
パラメータ	$k=\frac{1}{\beta}$, $t=-\frac{\alpha}{\beta}$	$k=\alpha$, $t=\beta$
標準誤差	$s(k)=\frac{s(\beta)}{\beta^2}$	$s(k)=s(\alpha)$
標準誤差	$s(t)=\frac{1}{\beta^2}\sqrt{\beta^2 s(\alpha)^2+\alpha^2 s(\beta)^2}$	$s(t)=s(\beta)$

(註)　x on y 回帰直線の傾きは$\beta=\frac{S_{yy}}{S_{xy}}$.

図－6　従来法と新しい方法におけるパラメータ数値を得るための数式集

表－1　数値例（$(DT)=k+tD$ 式と不偏長軸を用いた新しい方法の場合）

T	D	$X=D$	$Y=DT$	$(X_i-\bar{X})^2$	$(X_i-\bar{X})(Y_i-\bar{Y})$	$(Y_i-\bar{Y})^2$
15	58	58.0000	870.0000	1156.0000	14183.6667	174028.0278
17	30	30.0000	510.0000	36.0000	343.0000	3268.0278
20	19	19.0000	380.0000	25.0000	364.1667	5304.6944
23	15	15.0000	345.0000	81.0000	970.5000	11628.0278
26	12	12.0000	312.0000	144.0000	1690.0000	19834.0278
30	10	10.0000	300.0000	196.0000	2139.6667	23358.0278
Σ		144.0000	2717.0000	1638.0000 $=S_{xx}$	19691.0000 $=S_{xy}$	237420.8333 $=S_{yy}$
Mean		24.0000	452.8333			
n		6	6			

$\beta=t=12.0393$	$s(\beta)=s(t)=0.3287$	$r=0.9985$
$\alpha=k=163.8893$	$s(\alpha)=s(k)=9.5799$	$s'_e=13.3099$

お　わ　り　に

今回紹介した新しい数式と不偏長軸という line-fitting 法の組み合わせによる推定法は，従来の方法と比較して，データにバラツキのある場合ほど偏りの少ない推定値を与える。

読者の中には不偏長軸がもつ最尤値にまつわる数理統計学上の問題点を心配される向きがあるかもしれない。これについて筆者らは，個々のデータ点の平均発育日数とその標準誤差，それにデータ数についての情報がそろっていれば，数理統計学上このような問題のまったく生じない推定方法の開発に成功している（未発表）。これには複雑な試行プログラム計算が必要であり，代数的には解けないものなので，次の機会にご紹介したいと思う。

しかし今回紹介した方法による推定値は，上記プログラム計算により検証したところ，実用的にはまったく問題のない精度が得られている。安心してお使いいただきたい。

有効積算温度法則はシンプルであり，直感的にも受け入れやすく，長い間使われてきた。伊藤（1976）によると，18 世紀前半にフランス人レミュールが見いだした法則だとのことである。今現在も多くの研究者によって使われているのは，単純な法則の中に深い真理が隠されていると考えられるからであろう。筆者らもいろいろな昆虫についてのデータを集めて分析するうちに，種の分化にかかわる内容が含まれるように感じ始めている。さらに多くの生物種について，より正確なデータがどんどん蓄積され，この法則から発展したいろいろな仮説が打ち出されることを期待したい。

引　用　文　献

1) Cheng, C.-L. and J. W. Van Ness (1999) : Statistical Regression with Measurement Error, Arnold, London, pp. 262.
2) Ikemoto, T. and K. Takai (2000) : Environmental Entomology, **29** : 671～682.
3) 伊藤嘉昭 (1976) 動物生態学 (上巻), 古今書院, 東京, pp. 226.
4) Kermack and Haldane (1950) : Biometrika, **37** : 30～41.
5) Lamb (1992) : Environmental Entomology, **21** : 10～19.
6) Ricker, W. E. (1975) : J. Fish. Res. Board Can., **32** : 1494～1498.
7) Ricker, W. E. (1984) : Can. J. Zool., **62** : 1897～1905.
8) Solari, M. E. (1969) : J. Roy. Statist. Soc., Ser. B **31** (2) : 372～375.
9) スネデッカー・コクラン (1967) : 統計学的方法, 岩波書店, 東京, pp. 546.