

歯科疫学統計
ー第6報 傾向（トレンド）分析の基礎 その3ー
瀧口 徹

A review of oral epidemiological statistics
ー Part VI : The bases of Trend Analysis 3rd follow-up review ー
Toru Takiguchi

キーワード：トレンド（傾向）分析、形・非線形モデル、回帰直線の差、統計フリーソフト、統計R

はじめに

第4報¹⁾と第5報²⁾でトレンド（傾向）分析を扱ってきた。目的と手法が多義的で混乱しやすいトレンド（傾向）分析を便宜的に

1. 時系列データの不規則変動を除去して傾向の把握と予測性に重点を置いた分析
2. 最適回帰曲線算出に重点を置いた分析
3. 時系列データの類型化に重点を置いた分析
4. 量ー反応的要因分析に重点を置いた分析

の4種類に類型化して解説してきた。第6報では1の予測性と2の最適回帰曲線の算出とに重点を置いてトレンド（傾向）分析に関するこれまでのレビューを拡張・補完することにした。またその計算に適したフリーソフトである統計R (statistical R environment)^{Web01)}とMaxima 5.17.0^{Web02)}の特徴を合わせて紹介する。

1) 事例のプロフィール

1992年に行った中華人民共和国江西省南昌市に

おける成人592名の歯科疫学調査³⁾データから年齢と現在歯数、血圧、BMI (Body Mass Index)のデータを用い、比較対象として我が国の歯科疾患実態調査年のうち最近接の検査年である1993年の公表データを用いた。

2) 解説および解析手順

下記の手順で行う。

- (1) 統計フリーソフト活用の意義
- (2) 2つの回帰直線の同一性検定
- (3) 各種回帰モデルの係数の算出
- (4) 生物学的妥当性を考慮した最適回帰モデルの選定
- (5) モデルによる補内法、補外法による点予測

(1) 統計フリーソフト活用の意義

PC (パソコン) 用に市販されている統計ソフトパッケージに関してしばしば遭遇する問題やユーザの憂いを列記すると

- ①価格やライセンス取得の価格が高価で手に入りにくい。
- ②オーダメードの解析（プログラミング解析）ができない（しづらい）
- ③ソフトによって出来る解析と出来ない解析がある

【著者連絡先】

〒341-0003 埼玉県三郷市彦成3-86
深井保健科学研究所
主席研究員 瀧口 徹
TEL&FAX : 048-957-3315
E-mail : taki8020@mth.biglobe.ne.jp

- ④一般に出力情報(統計量)が多数(過剰)で、一般的でない統計量でも統計的意味合いの解釈を専門書に委ねている場合が多いので消化不良に陥りやすい
- ⑤ソフト間で解析法や出力される統計量が同質異名、または同名異質の場合があり解釈の誤りを引き起こす場合がある
- ⑥同一ソフトでの解析でも解析のオプションのパラメータ選択に際して困惑する
- ⑦計算精度の信頼性が(ときに)不明
- ⑧グラフ作成機能に制限が多い(かゆいところに手が届かない)
- ⑨統計ソフト間でのデータファイルの互換性(インターフェース)の問題が生じやすい(コピーペーストしづらい)
- ⑩各種分析を効率よく行う外部のライブラリー(プログラムパッケージ)が利用できない(しづらい)

等であろう。①のコストに関連してフリーソフトの利点については説明を要しない。しかし単に個人が市販ソフトを購入可能かの問題ではなく、高額なライセンス登録費用を負担して集団を対象として統計の実践的教育・訓練をしている場合には質の高いフリーソフトは十分一考に値するソフトであろう。②に関連しては、統計量と数値計算のアルゴリズム等がわかればフォートランや各種プログラム言語で重回帰分析、ロジスティック分析、因子分析等いずれも100%のステップが作製可能である。しかし実際にはそうしたことは専門の研究者でないと不可能に近い。それで実際には基礎統計量の算出も含め各種検定、多変量解析等の複雑な統計処理は統計ソフトを利用することになるのが必然である。ただユーザとしては、欲を云えばもう少し中間的なプログラム言語で、例えば関数や統計ライブラリーを組み合わせて使用し、1つのまとまった統計処理に仕上げるソフトのニーズも高いと思われる。SASがその高度なソフトの代表、Excelがいわば一般のソフトの代表といえるが、今回紹介する統計RとMAXIMAもまた中間プログラミングが可能なソフトである。

Excelのように身近でかつSASのように高度という特徴がある。③に関連しては、ある1つのプログラムパッケージにある解析法または統計量がA)未掲載のため計算が不可能な場合、B)マニュアル計算をすれば可能な場合(中間統計量により)、の2通りがある。経験的にA)不可能な場合、が多い解析法はFisherの直接確率法、Cochran-Armitage-trend検定、Mantel-Haenszel検定、Mantel extension trend検定、4次元以上の回帰式の推定、S字状曲線やロジスティック曲線以外の非線形回帰式の推定、異なる集団の回帰直線の差の検定、各種乱数を用いたシミュレーション(モンテカルロ法等)等である。ここで興味深い資料^{Web03)}を紹介したい。この資料は15種類の市販の代表的統計ソフトについて695種類の統計分析法または最終統計量のどれが可能かの一覧である。例えばSPSS 10.0とStatXact 5とは取載率が双璧であり最も多くの分析法または統計量をカバーしている。しかし、いずれもそれぞれ695種類の約25%をカバーしているに過ぎない。このことは単一の統計ソフトで万事片付かず起承転結出来ない場合が往々にしてあることを如実に示している。次にB)マニュアル計算で可能な場合、に関しては遭遇する問題点④の出力される中間統計量の多さと裏腹の関係にある。例えば回帰直線の傾きの値だけが出力されて変数 x 、 y の偏差平方和： SS_{xy} と変数 x の偏差平方和： SS_x が出力されない今回の解説および解析手順(2)の2つの回帰直線の傾きの検定が不可能である。しかし回帰直線の傾きは変数 x 、 y の偏差平方和/ x の偏差平方和($\text{slope} = SS_{xy} / SS_x$)で計算されるので、逆に傾きと分散と例数によって SS_{xy} と SS_x をマニュアルで算出し、算定式に当てはめれば手間がかかるものの2つの回帰直線の傾きの検定が可能ではある。2つ以上の回帰直線の傾きの差検定は共分散分析(ANCOVA)^{Web04)}の交互作用項を使っても判定できるが、2つの回帰直線傾きと切片のt検定が用意されている統計ソフトは前述の15種の統計ソフト比較表^{Web03)}によるとStatEasy^{Web05)}だけである。

ここで⑤に関連して同一名の解析法でアルゴリズムが異なる例としてSPSSの尤度比による変数選択法⁴⁾が挙げられる。このSPSSの解析法の優れた点は、各ステップで全ての変数一つずつ削除したモデルを全て比較した上で最も「-2対数尤度の変化 (=deviance差)」が小さくなる変数を削除している点である。つまり、いわば石橋を叩いて渡るアルゴリズムを採用している点で他の解析ソフトと同一名の分析でも結果が違って来る場合が当然出てくる。また統計量の同質異名に関して、SPSSの日本語マニュアルの「-2対数尤度」の標記は「-2最大対数尤度： $-2 \times \log \text{likelihood}$ 」を示しSTATAのGRIM分析⁵⁾のディビアンズ (deviance) と全く同じ統計指標である。従ってSPSSの「-2対数尤度の変化」による方法はSTATAのディビアンズ差による方法と呼び方が違うだけで全く同じである。また同名異質に関してはSTATAのAICは例数nで除したAIC/nで標記される厳密には別の統計量であり勘違いしやすいので注意を要する。また統計RのAIC関数はパラメータk=2とすると赤池が開発したAIC (Akaike's Informa-

tion Criterion) が $k = \log(n)$ (n: 例数) とするとShwarzが開発したBIC (Bayesian Information Criterion) となる。余談になるがこの統計Rの仕切りはAICとBICのいずれが妥当かの1978年来の長年に渡る論議^{6,7)}の整理法として、いずれか片方とか並列して出力するのではなくBICをAICの変形として位置づけるという意味で絶妙の取り扱いである。⑥のオプションによる結果の違いはどのオプションが最適かの判断を個別にしなければならぬ。例えば多重ロジスティック回帰分析 (MLRA) の変数減少法を尤度比⁵⁾による方法 (=deviance差による方法) を選択するかWald⁵⁾による方法によるかは後者は前者の代用⁴⁾なので算出されるなら前者が望ましい。またカテゴリーデータの対比 (contrast) を単純 (“simple contrasts”) とし最初のカテゴリーを参照カテゴリーとするか、あるいは対比を指標 (“indicator contrasts”) とし対象カテゴリーを除いた全体を参照カテゴリーとするかはカテゴリーデータの性質による。次に⑦の計算精度に関してはテストデータを用いて、例えばスパコン使用のSASの演算結果をゴールドスタンダードとして比較検証すればい

表1 統計Rの特徴

1.	統計Rはベル研究所のJohn Chambersらによって1970年代の始めに開発され、S、S-PlusとS言語を共通言語とし、アルゴリズムは1990年代の初めニュージーランドのWaikato大学の Ross Ihaka、Robert Gentlemanらによって、S、S-Plusを見本にしてフリー統計システムとして開発された。 http://www.daa.com.au/analytical-ideas/databa
2.	英名はThe statistical R (program, environment) または単純にRとして表現されており、和名は「統計R」または「R」として表現されている例が多い。しかし単純にRだと(重)相関係数と間違えたり半徑と紛らわしのでこれらを選けるためにはやはり「統計R」と表現するのが望ましいと思われる。
3.	the General Public License (GPL)でフリーソフトとしてダウンロード出来るために、SAS、SPSS、S-PLUS等の商用統計パッケージを購入する余裕のない学生等のユーザーを含め統計学教育と解析の実践に際して国際的な標準語としての役割を果たしてきている。 ダウンロード： http://www.r-project.org/ 各国の研究機関のミラーサイトから選択 (ミラーサイト例)筑波大学 http://cran.md.tsukuba.ac.jp/bin/windows/base/release.htm
4.	多くの統計学者、S-plusを始めとする統計ソフトの開発者、研究者(利用者)が改善に参画し続けているため、解析法の質、量とも進化し続けている。
5.	EXCEL、SPSS、STATA等の統計ソフトとのインタフェイスがありデータ互換が容易である。
6.	多数(87個 2009/8 時点)のプログラムパッケージと関数が用意されており簡単に利用できる。 例 nlme:線形非線形混合モデル解析 関数: fisher.test(): フィッシャーの直接確率法
7.	線形、非線形関数モデル、混合モデル等のマニュアル計算が容易である。
8.	S-Plusに近似したプログラム解析によりオーダーメードの解析ができる。 例 モンテカルロ法の適用
9.	高度なグラフィックス機能がありパワーポイントファイル等へのコピー・ペーストが簡単である。
10.	R Commander(日本語版)のインストールにより、SPSS的な解析選択画面で解析ができる。
11.	Rのセットアップ+R入門 http://cwoweb2.bai.ne.jp/~jgb11101/files/cart/cart.html

いのだが多くのPCユーザにとって实际的でない。しかし非線形式の解をニュートン法などで求める場合にはソフト間で結果の違いが出やすい。いずれにせよ市販ソフトの精度が正しくてフリーソフトの精度は問題があると単純には片付けられない。今回紹介する統計RやMAXIMAの場合は作製主体（組織）の信頼性とフリーソフト故に国際的なフィードバック下での膨大な使用頻度に信頼性の補償を置いているといえるであろう。⑦～⑩については⑨の外部ファイルとの互換性、⑩外部ライブラリーの利用が重要である。

ここで上記①～⑩の「統計ソフトの憂い」をかなり解消する方法として表1にフリー統計ソフトの統計Rの特徴を示した。統計Rは上記①、②、④、⑦、⑧、⑨、⑩を概ね満たす。本統計はS-Plusといわば兄弟のフリーソフトであり1990年代に国際的な統計フリーソフトとしての開発プロジェクトを開始した。EXCEL等の表計算ソフト（アドイン統計ソフトを含む）、SPSSやSTATAなどの統計プログラムパッケージでカバーしていない解析や統計量算出が可能で利便性の高いソフトである。パッケージはR Console画面で

“library ()” でリストを確認できる。ダウンロード法やマニュアルは表1中の引用サイトを参照されたい。加えてRコマンドというライブラリーをインストールすることにより図1に示す日本語のダイアログボックス（メニュー選択画面）が提供されSPSSやSTATA等に類似したダイアログボックスが利用できる。また、表2に数値計算フリーソフトMAXIMAを紹介した。MAXIMAは①、②、⑦を満たす優れたものである。

(2) 2つの回帰直線の同一性検定 (統計R, (Excel))

図2に2つの回帰直線の類似性の分類を示す。2つの有意な回帰直線の傾きと切片の関係^{Web06)}はタイプ1：同一（傾き、切片とも有意差なし）、タイプ2：平行（傾き有意差なし、切片有意差あり）、タイプ3：傾き有意差あり、切片有意差なし、タイプ4：不一致（傾き、切片とも有意差あり）の4型である。独立変数： x_1 、と従属変数： y_1 、同様に x_2 と y_2 の関係、それぞれに回帰直線の傾きと切片の検定、あるいはコ克蘭・アーミテージ検定等のトレンド検定^{1,2)}が可能である。これらトレンド検定と回帰直線の同一性の検定との違いは



図1 統計Rのダイアログボックス (R コマンドー)

表2 数値計算フリーソフト MAXIMA5.17.0

1. Maxima はGNU General Public License (GPL) 準拠で配布されている数式処理ソフト。米エネルギー省(DOE)や米航空宇宙局(NASA)等の資金協力を得ながらマサチューセッツ工科大学(MIT)で開発され、シンボリックな計算ができる強力な汎用数式処理ソフトである。
2. 本報では年齢と現在歯数の非線形関数モデル式(ロジステック関数)を用いて特定の現在歯数に到達する年齢の推定を行った。
3. 関連Web-site
Download
<http://sourceforge.net/projects/maxima/files/Maxima/5.9.3-Windows/maxima-5.9.3.exe/download>
日本語マニュアル
http://www.bekkoame.ne.jp/~ponpoko/Math/maxima/maxima_toc.htm

何であろうか。回帰直線の同一性の検定は2つのトレンドの差の検定である。事例としてはアフリカの HIV (+) と HIV (-) の女性の出産後の体重線形変化⁸⁾の差の検定、風疹抗体価の比較⁹⁾、唾液の歯牙表面における連鎖球菌に対する作用比較¹⁰⁾等医歯学の分野においても多数の適用例がある。

ここで今回の事例を図3に示す。これは中華人民共和国江西省南昌市における1992年の歯科疫学

調査¹⁾のデータから20歳~80歳未満の男女計592名の年齢と現在歯数およびDMFTの散布図および男女それぞれの回帰直線である。この調査の対象者は60歳になるまで現在歯数が驚異的多いため日本では明確な年齢とともに現在歯数が右肩下がりになる傾向が視覚的には余りはっきりしないのは驚くべき事である。しかし本調査は同地区の労働者のランダムサンプルなので地域代表として偏っているわけではない。回帰直線の男女差は傾

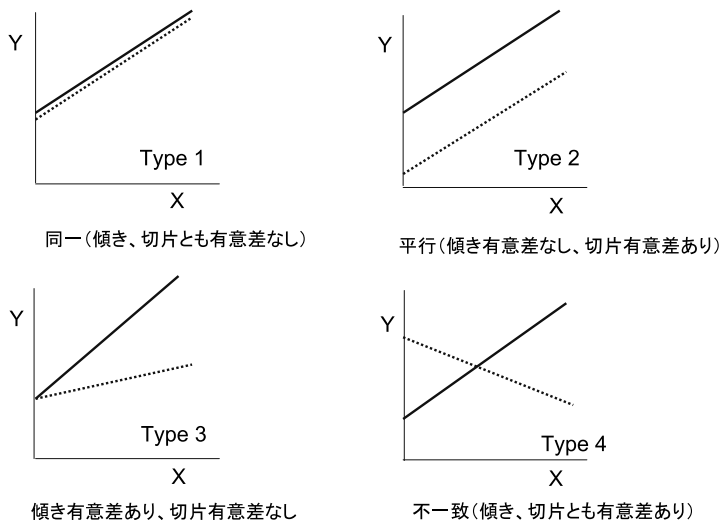
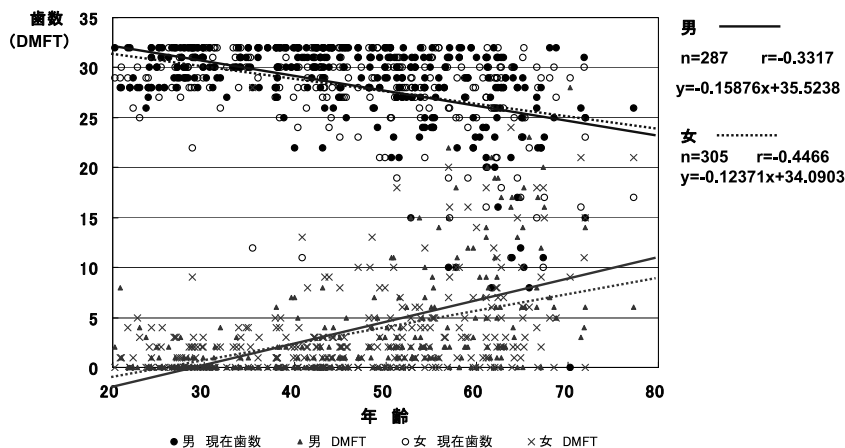


図2 2つの直線の傾きと切片の関係の4型



結果: 回帰直線の同一性検定 現在歯数 傾き: $p=0.2094$ 切片: $p=0.2076$
DMFT 傾き: $p=0.4853$ 切片: $p=0.3693$

図3 中華人民共和国成人の年齢と現在歯数およびDMFTの性差 (散布図と回帰直線)

きと切片の差 (0歳時点) それぞれの統計学的有意性で判定される。本図をみると視覚的には若干の性差があるように見えるが果たして統計的に有意であろうか。表3に統計Rでの分析手順と結果を示す。計算ステップが煩雑に見えるが近似した統計指標と中間指標が繰り返してくるので見かけほど煩雑ではない。解析の結果は有意確率が傾きの差で $p=0.2094$ 切片の差で $p=0.4853$ といずれも有意で無い。従って男女の回帰直線は差が有意でないという意味で同一 (タイプ1) に同定される。また同じく図3に示す年齢とDMFTの回帰直線も傾き: $p=0.4853$ 切片が $p=0.3693$ で性差がなくタイプ1の関係である。回帰直線の差が有意な例として同じ中国のデータから図4に年齢と最高血圧 (SBP) の例を示す。回帰直線の同一性検定で傾きが $p=0.0315$ 、切片が $p=0.2687$ で男性に比して女性の最高血圧の加齢による上昇傾向が有意に急勾配である。一方、最低血圧 (DBP) は傾きではなく、切片の差が有意であった。しかしBMIの回帰直線性差は有意でなく、肥満度の血圧の性差への影響は少ないことが示唆された。現在歯数、DMFT、最高血圧の例でわかるように見かけでは差が有意かどうか分からない場合でこうしたトレンドの差の検定は意義がある。

なお、回帰直線の同一性の検定は手順が多少煩雑であるが表計算ソフト (Excel) で積み木細工の要領で計算結果を積み上げていけば計算が可能である。

(3) 各種回帰モデルの係数の算出 (統計R)

表4に統計Rによる最適回帰モデルの選定と図示のためのコマンドを示した。統計Rのダウンロードは表1に示したhttpを、コマンドと操作法の詳細はウェブサイド^{Web01}) またはガイドブック¹¹⁾等を参照されたい。表4では紙面の関係で1次、2次、6次、S字、ロジスティックの各関数モデル¹²⁾のみをデモとして扱っている。係数とAIC計算、および年齢と現在歯数の関係を図示するためのコマンドであり、図5はそのデモで男性287名の出力の結果である。年齢を示すx軸をplot () 関数で120歳まで拡張し、predict () 関数で各関数による予測値を求めlines () 関数で軌跡を図示している。表5において線形8種、非線形4種計12種のモデルの適合度を性別にAIC^{1,2,4,5)}でそれぞれの回帰線の適合度にランクを付けている。それによると男性では2次、3次、4次、5次、ロジスティック式の順に適合が良く、女性では3次、4次、6次、2次、5次式の順であった。

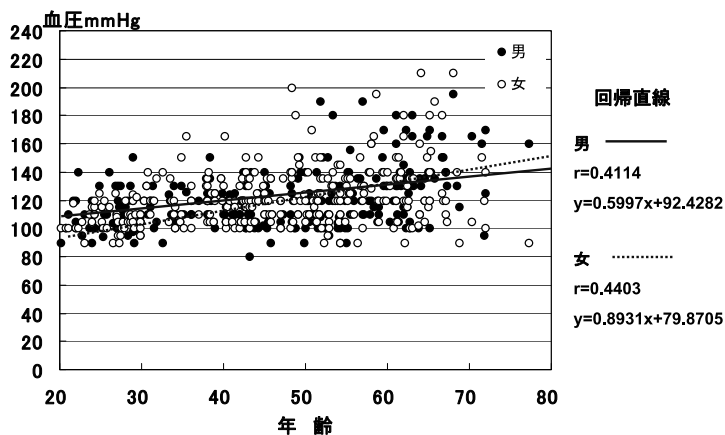
表3 2つの回帰直線の同一性（傾き、切片）の判定 ー統計Rによる解析ー

```
# 男性 287 人の SPSS データ入力
> Dataset <- read.spss("L:/Data/China/China data male 287.sav", use.value.labels=TRUE,
max.value.labels=Inf, to.data.frame=TRUE)
# 男性データの直線回帰式の傾き切片その他差の検定に必要な統計量の計算
> x <- Dataset$AGE > y <- Dataset$PRESENT
> N1 <- length(x) → 287
> SIGX1 <- sum(x) > SIGY1 <- sum(y)
> MEANX1 <- mean(x) → 46.22718
> MEANY1 <- mean(y) → 28.18467
> SSX1 <- sum(x^2)-sum(x)^2/N1 → 52412.33
> SSY1 <- sum(y^2)-sum(y)^2/N1 → 6623.213
> SSXY1 <- sum(x*y)-sum(x)*sum(y)/N1 → 8321.14
> SLOPE1 <- SSXY1/SSX1 → 0.1587630
> INTERCEPT1 <- (SIGY1 -SLOPE1*SIGX1)/N1 → 35.52384
> RSS1 <- SSY1-(SSXY1)^2/SSX1 → 5302.123
> RDF1 <- N1-2 → 285
# 女性 305 人の SPSS データ入力
> Dataset <- read.spss("L:/Data/China/China data female 305.sav", use.value.labels=TRUE,
max.value.labels=Inf, to.data.frame=TRUE)
# 女性データの直線回帰式の傾き切片その他差の検定に必要な統計量の計算
上記男性と同じ内容の計算を実施（計算手順省略）
> SLOPE2 → -0.1237114
> INTERCEPT2 → 34.09028
# 回帰線の傾きの差の検定に必要な共通統計量（共通残差分散等）の計算と t 検定
> SYXP12 <- (RSS1+RSS2)/(RDF1+RDF2) → 16.56936
> Sb1b2 <- sqrt(SYXP12/SSX1+SYXP12/SSX2) → 0.02789148
> tvalueS <- (SLOPE1-SLOPE2)/Sb1b2 → t=-1.256715
> probS <- pt(abs(tvalueS), df=N1+N2-4, ncp=0, log = FALSE) → 0.895322
> probS2tail <- 2*(1-probS) → p=0.2093559 有意差無し
# 回帰線の切片の差の検定に必要な共通統計量（共通残差分散等）の計算と t 検定
> Ac <- SSX1+SSX2 > Bc <- SSXY1+SSXY2 > Cc <- SSY1+SSY2 > SSc <- Cc-Bc^2/Ac
> DFc <- N1+N2-3 > RMS <- SSc/DFc > bc <- Bc/Ac
> bunshi <- (MEANY1-MEANY2)-bc*(MEANX1-MEANX2)
> bunbo <- sqrt(RMS*(1/N1+1/N2+(MEANX1-MEANX2)^2/Ac))
> tvalueI <- bunshi/bunbo → t=-0.6983008
> probI <- pt(abs(tvalueI), df=N1+N2-4, ncp=0, log = FALSE) → 0.7573674
> probI2tail <- 2*(1-probI) → p=0.4852652 有意差なし
Reference: http://www.hku.hk/ecology/staffhp/kl/Biom12.ppt
```

(4) 生物学的妥当性を考慮した最適回帰モデルの選定（統計R）

12種類の回帰式の生物学的妥当性（整合性）(biological plausibility)¹³⁾の判定は今回、各曲線のデータ値外（80歳以上）の軌跡の非論理性と男女間の傾向の非論理性で判定した。例えば図6, 7に示すように6次曲線は男女とも70歳前後までは適合度がいい軌跡を示しているがこれを過ぎると男

性は急上昇、女性は急降下するという異常な軌跡を呈し男女間の差が説明できない非論理性を示した。このため6次式を生物学的妥当性に欠けると判断した。3次式から5次式も同様な理由で生物学的妥当性が欠けると判断した。一口に生物学的妥当性といっても各種生物学的指標が例えばS字曲線に近似したパターンで変化するわけではない。判断は時に恣意的になりやすく、新しい知見



結果: 回帰直線の同一性検定 傾き: $p=0.0315$ 切片: $p=0.2687$

図4 中華人民共和国成人の年齢と最高血圧の性差 (散布図と回帰直線)
 - 対象者: 中華人民共和国江西省南昌市成人591名 1992年 -
 (注) データ値欠損のため図3より対象者数が1名少ない

を排除してしまう点を注意する必要¹³⁾がある。多項式は内挿、補間 (interpolation) の条件では適合度が高いが外挿、補外 (extrapolation) では異常な軌跡を呈しやすい。一方、上に凸の2次曲線は20歳未満の歯列成長期には現在歯数が少なくなることを捕らえればその軌跡の生物学的妥当性が保たれる。軌跡は年齢とともに急降下するが生物学的妥当性を欠くとまでは言えない。外挿、補外の条件では3~6次式が生物学的妥当性を著しく欠く軌跡を示すことから、3~6次式を除外した8種類の回帰式から最適回帰式を選択することとした。

一方、直線式はロジスティック曲線と比較してデータの両端の部分において適合度が下がるが適合度が3番目であり年齢と現在歯数間の相関係数: r は男性が -0.45 ($p < 0.001$ 、女性が -0.33 ($p < 0.001$), かつ回帰直線の傾きが高度に有意で、年度間、男女間の傾向の比較に有用であることが判明した。

(5) モデルによる補内、補外法による点予測 (統計R, MAXIMA)

ここで統計モデルの類型化でよく出てくる線形

モデルと非線形モデルの違いについて表6に示した。モデルの関数を (偏) 微分し一次導関数を求めたとき元のパラメータが残る場合を非線形モデルという。べき乗 (a^x , e^x) を関数に含んでいる成長曲線モデル等がこれに当たる。最小二乗法を用いた非線形モデルの係数を求めること、および非線形モデルの方程式の解は一般に解析的には求まらない。そこでニュートン法 (Newton-Raphson法)⁵⁾ 等によって精度のいい近似解を得る。表7にAICによる適合度判定の上位3位の線形 (1次式、2次式)、非線形モデル (ロジスティック式) による日本と中国の年齢 (x軸) と現在歯数 (y軸) の点予測値を示す。このうち男性の解析結果を図8に示し、AIC判定により最適の二次式とロジスティック式の軌跡と年齢および現在歯数の点予測値の関係を示している。すなわちx軸は年齢を示し100歳時点において二次式とロジスティック曲線ではそれぞれy軸の現在歯数が幾つになるかを示している。データの存在する範囲内の予測である内挿法における予測は二次式、ロジスティック式の間で大きな違いはみられないがデータの存在しない領域の予測である外挿法の予測値は大きく異なることが見てとれる。このように、

表4 統計Rによる最適回帰モデルの選定と図示

```

Dataset <- read.spss("L:/data/china/China data.sav", use.value.labels=TRUE,
max.value.labels=Inf, to.data.frame=TRUE) ※SPSS ファイルからのデータ読み込み
x <- Dataset$AGE ※変数 AGE を x とする
y <- Dataset$PRESENT ※変数 Present を y とする
numSummary(y) ※yの頻度および頻度% 表示

x1 <- x
x2 <- x^2
x3 <- x^3
x4 <- x^4
x5 <- x^5
x6 <- x^6 } ※x 1・x 6 に変数 x のべき乗値を挿入

Model.L1 <- lm(y~ x1)
Model.L2 <- lm(y~ x1 + x2)
Model.L6 <- lm(y~ x1 + x2 + x3 +x4 +x5 +x6)
Model.S
nls(y~exp(a+I(b/x1)),start=list(a=0.1,b=1),trace=TRUE)
bun <-32 ※ロジステック関数の上限値を 32 とする
Model.LGSbun <-
nls(y~1/(1/bun+a*(b^x1)),start=list(a=0.01,b=1.00),trace=
TRUE)

print(summary(Model.L1))
print(summary(Model.L2))
print(summary(Model.L6))
print(summary(Model.S))
print(summary(Model.LGSbun))
cat("AIC of Model.L1=", AIC(Model.L1),"¥n")
cat("AIC of Model.L2=", AIC(Model.L2),"¥n")
cat("AIC of Model.L6=", AIC(Model.L6),"¥n")
cat("AIC of Model.S=", AIC(Model.S),"¥n")
cat("AIC of Model.LGSbun=", AIC(Model.LGSbun),"¥n")
plot(x,y,xlim=c(20,120),ylim=c(0,35),main="年齢と現在歯数",xlab="年 齢",ylab="現在歯
数",cex=1,col=8,pch=20)
legend(20,10,c("一次式","二次式","六次式","S 字","ロジステック"),bty=0,box.lty=0,
lwd=2,col=c(8,8,8,1,1),lty=c(1,2,3,4,1))
xv <- seq(20,120,1.0)
yv <- predict(Model.L1,list(x1=xv))
lines(xv,yv,lty=1,lwd=2,col=8)
yv <- predict(Model.L2,list(x1=xv,x2=xv^2))
lines(xv,yv,lty=2,lwd=2,col=8)
yv <- predict(Model.L6,list(x1=xv,x2=xv^2,x3=xv^3,x4=xv^4,x5=xv^5,x6=xv^6))
lines(xv,yv,lty=3,lwd=2,col=8)
yv <- predict(Model.S,list(x1=xv))
lines(xv,yv,lty=4,lwd=2,col=1)
yv <- predict(Model.LGSbun,list(x1=xv))
lines(xv,yv,lty=1,lwd=2,col=1)
    
```

※1次関数、2次関数、6次関数、ジグモイド関数、ロジスティック関数の係数計算

※1次関数、2次関数、6次関数、ジグモイド関数、ロジスティック関数の係数計算結果を出力

※1次関数、2次関数、6次関数、ジグモイド関数、ロジスティック関数のAIC計算

※作図

特に外挿法による予測に際しては単に数学的、統計的なモデルではなく生物学的妥当性等の判断を入れてモデルを選択することが重要である。また現在歯数20歯（y軸）が2つの回帰式でそれぞれ何歳になるかを示している。また表7に示すように逆に20歯および10歯に到達する年齢をMaxima^{Web02}でニュートン法を用いて計算した。

ロジスティック回帰曲線は回帰直線と比較して

2つの曲線の違いを定量的に表現するのが難しいが、今回の事例のようにAIC等を用いた適合度比較で回帰直線より精度がいい場合は内挿法、外挿法の予測に有用である。変曲点はロジスティック関数の2次導関数（2回（偏）微分したもの）が0になる点であり、-そのx、y座標は（-Ln（au）/Ln（b），u/2）となる。日中間、男女間の変曲点の違いに注目すると中国で10年以上の差で女性の

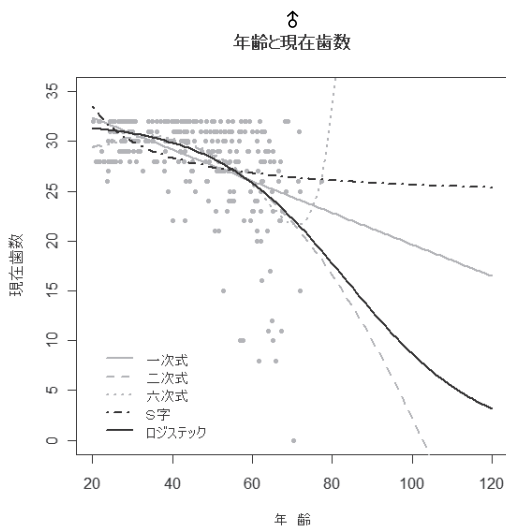


図5 出力結果デモ (例 男 287名)

表5 線形、非線形関数による年齢と現在歯数関係モデル - データ：中華人民共和国江西省南昌市 1992 -

モデルの妥当性の判断

Linear/Non-linear 線形、非線形区分	Model	和名	統計的(数学的)適合性				統計的適合性+生物学的妥当性			
			男性 AIC	男性 順位	女性 AIC	女性 順位	Biological plausibility 生物学的妥当 性	男性 順位	女性 順位	
Linear Model 線形モデル	Linear	一次式(線形)	1657.5	7	1688.4	7	○	3	3	
	Quadratic	二次式	1638.8	1	1679.3	4	○	1	1	
	Cubic	三次式	1639.4	2	1677.2	1	×			
	Quartic	四次式	1639.5	3	1679.0	2	×			
	Quintic	五次式	1640.3	4	1679.8	5	×			
	Sextic	六次式	1641.8	6	1679.1	3	×			
	Inverse	逆数	1683.2	11	1704.6	11	○	7	7	
	Logarithmic	対数	1670.1	9	1696.3	9	○	5	5	
Non-Linear Model 非線形モデル	Sigmoid	S字曲線	1685.4	12	1705.6	12	○	8	8	
	Power	べき乗	1672.8	10	1697.6	10	○	6	6	
	Exponential	指数	1660.3	8	1689.7	8	○	4	4	
	Logistic	ロジスティック	1641.4	5	1681.2	6	○	2	2	

変曲点が高齢であるのに対し、日本では5年弱の差で逆に男性の変曲点が高齢である点は興味深い。

おわりに

今回は「歯科疫学統計」シリーズの第6報、またトレンド分析の第3報として特にフリーソフト

である統計Rの活用の意義と実際例を紹介した。誤解ないようにしておきたいことは、単に経済的理由により「フリーソフト」を推奨しているのではなく、重視しているのは次の2つの点である。まず1点目は既存の統計ソフトによる解析を身近にある統計RやMAXIMAなどのフリーソフトによって補完することにより解析の効率化が図られ

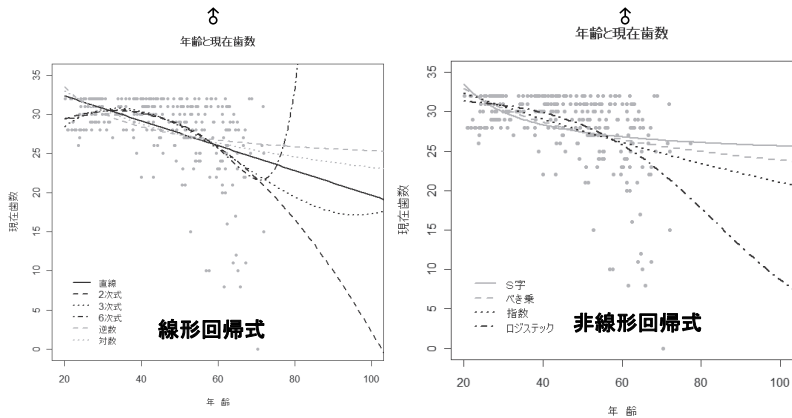


図6 中華人民共和国成人の年齢と現在歯数の回帰線（男 287名）

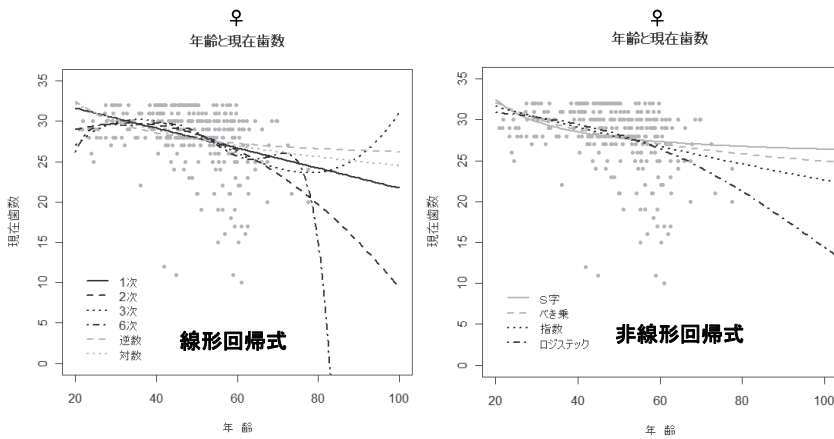


図7 中華人民共和国成人の年齢と現在歯数の回帰線（女 307名）

表6 線形モデルと非線形モデル

モデルを各パラメータで偏微分したときに、**一次偏導関数がパラメータ自身を含むS字曲線やロジスティック曲線モデル等を非線形モデルと呼ぶ。**

例えば、

関数モデル	一次導関数	判別	分類
1次式: $y=a+bx$			
2次式: $y=a+bx+cx^2+dx^3$	$y'=b+2cx+3dx^2$	含まない	線形
対数式: $y=a+b \cdot \log_e x$	$y'=b/x$	含まない	線形
S次式: $y=\exp(a+b/x)$	$y'=(-b/x^2) \cdot \exp(a+b/x)$	$\exp(a+b/x)$ を含む	非線形
ロジスティック式 ¹⁾ $y=(1/u+ab^x)^{-1}$	$y'=-ab^x \cdot \text{LN}(b) / (1/u+ab^x)^2$	$(1/u+ab^x)$ を含む	非線形

1) uは最大値 式の出典はSPSSガイド : <http://www.washington.edu-Base UserGuide17.0.pdf>

表7 線形、非線形モデルによる年齢と現在歯数の予測 - データ：中華人民共和国（1992年）と日本（1993年）の比較 -

中華人民共和国江西省 1992						予 測				
種類	AIC 順位	係数(パラメータ) ^{a)}			変曲点 ^{b)} (歳)	80歳現在 歯数	100歳現 在歯数	20歳到達 年齢 ^{c)}	10歳到達 年齢 ^{c)}	
		b ₀ (切片)	b ₁	b ₂						u (極大値)
男性	一次式(線形) ^{d)}	3	35.52384	-0.15876	×	22.8	19.6	97.8	160.8	
	二次式 ^{e)}	1	23.61790	0.41341	-0.00627	×	16.6	2.3	73.8	90.1
	ロジスティック ^{f)}	2	0.0001985	1.06200		32	84.10	18.0	8.9	75.6
女性	一次式(線形)	3	34.09028	-0.12371	×	24.2	21.7	113.9	194.7	
	二次式	1	24.86549	0.29654	-0.00452	×	19.7	9.3	79.2	98.9
	ロジスティック	2	0.0004732	1.04487		32	95.47	21.2	14.4	83.8

日本歯科疾患実態調査 1993						予 測				
種類	AIC 順位	係数(パラメータ)			変曲点 ^{b)} (歳)	80歳現在 歯数	100歳現 在歯数	20歳到達 年齢 ^{c)}	10歳到達 年齢 ^{c)}	
		b ₀ (切片)	b ₁	b ₂						u (極大値)
男性	一次式(線形)	3	38.81179	-0.35283	×	10.6	3.5	53.3	81.7	
	二次式	2	24.90405	0.3140423	-0.00666	×	7.4	-10.3	59.5	76.4
	ロジスティック	1	0.0001162	1.08800		30	67.10	7.6	1.8	58.9
女性	一次式(線形)	3	39.60930	-0.44260	×	4.2	-4.7	44.3	66.9	
	二次式	2	26.73259	0.22064	-0.00618	×	4.8	-13.0	55.4	72.9
	ロジスティック	1	0.0001172	1.09400		30	62.89	5.3	1.0	55.2

注 a) 統計法R (Statistical R Environment) によって算出
 b) 変曲点: 凹凸の変化点 ロジスティック曲線の二次導関数=0の点 x軸: 変曲点年齢=-(LN(b₀+LN(u))/LN(b₁)) y軸: u/2
 c) 演算ソフトMaxima 5.17.0 によって計算
 d) $y = b_0 + b_1 * x$ x: age
 e) $y = b_0 + b_1 * x + b_2 * x^2$
 f) $y = 1 / (1 / u + b_0 * b_1 * x^u)$ u: 最大値 出典: <http://www.washington.edu/Base/UserGuide/17.0.pdf/>

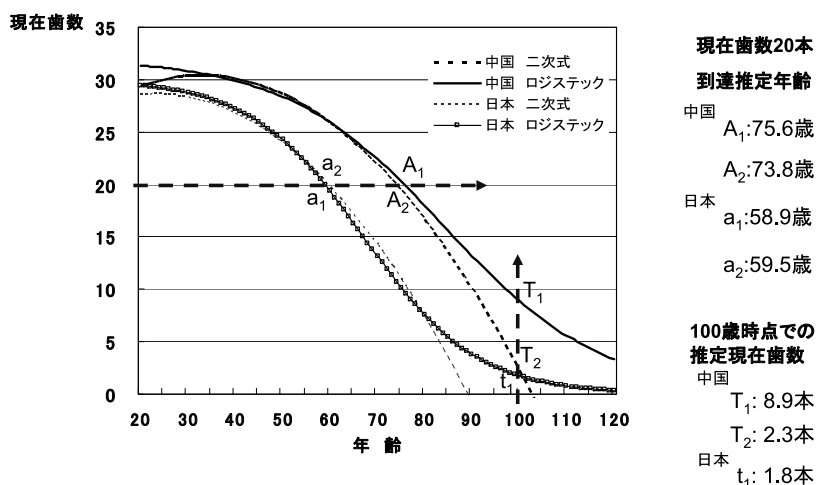


図8 二次曲線、ロジスティック曲線モデルによる予測 (点推定) - 男 -

解析結果に深みが出るという点である。次に、近未来の環境変化への備えである。インターネットの普及により情報の量のみならず質の担保が進化している。例えばいわば一つの財産であった百科事典や高価な国語や外国語の辞書と同質の情報が

インターネットで簡単に、しかもそれ自体は無料で入手できるようになった。これらの現象は益々加速し手元に百科事典や辞書を置く必要はほとんど無くなっていくであろう。少なくとも技術的には今でも十分可能である。問題は情報提供者の思

表8 MAXIMA 5.17.0を用いたニュートン法による非線形方程式 (logistic) の解法 ー現在歯数20、10本時の年齢は何歳かー 中華人民共和国江西省 (1992) の場合

```

Maxima 5.17.0 http://maxima.sourceforge.net
Using Lisp GNU Common Lisp (GCL) GCL 2.6.8 (aka GCL)
Distributed under the GNU Public License. See the file COPYING.
Dedicated to the memory of William Schelter.
The function bug_report() provides bug reporting information.
(%i1) load(newton1);
(%o1) C:/PROGRA~1/MAXIMA~1.0/share/maxima/5.17.0/share/numeric/newton1.mac
現在歯数=20 (%i2) k:20;x0:60;newton(1/(1/32+0.0001985*1.06200^x)-k,x,x0,1/10000);
(%o2) 20
<男> (%i3) (%o3) 60
(%i4) (%o4) 推定年齢 75.60876614877923
現在歯数=10 (%i5) k:10;x0:60;newton(1/(1/32+0.0001985*1.06200^x)-k,x,x0,1/10000);
(%o5) 10
(%i6) (%o6) 60
(%i7) (%o7) 推定年齢 97.20800640555528
-----
現在歯数=20 (%i8) k:20;x0:60;newton(1/(1/32+0.0004732*1.04487^x)-k,x,x0,1/10000);
(%o8) 20
<女> (%i9) (%o9) 60
(%i10) (%o10) 推定年齢 83.82847784450503
現在歯数=10 (%i11) k:10;x0:60;newton(1/(1/32+0.0004732*1.04487^x)-k,x,x0,1/10000);
(%o11) 10
(%i12) (%o12) 60
(%i13) (%o13) 推定年齢 113.4297800977362
(%i14)
    
```

<http://maxima.sourceforge.net/docs/intromax/intromax.html>

想、知的所有権の扱いおよび採算性であろう。これと同様、フリーの統計ソフトの代表である統計Rも世界各国の膨大なユーザの力で進化している。高価なライセンス料で手に入れた統計ソフトCDからPCにインストールして特殊で高尚な解析をするという手立てもそう遠くないうちに相当変わるであろう。その時のための予行演習として本報が役立てば幸いである。

謝辞：本総説を執筆するにあたり各段階でご指導いただいた歯科疫学研究会の顧問である重松逸造、蓑輪眞澄、高江洲義矩、境 脩、伊藤学而、佐々木英忠の各先生に深謝申し上げます。さらに貴重なご助言をいただいた深井稜博同研究会会長をはじめとする幹事の方々に感謝申し上げます。

参照 Web-site (2009/8/16 現在)

Web01) <http://cwoweb2.bai.ne.jp/~jgb11101/files/cart/cart.html> : Rのセットアップ+R入門

Web02) http://www.bekkoame.ne.jp/~ponpoko/Math/maxima/maxima_toc.html : MAXIMA 日本語マニュアル

Web03) http://docs.google.com/gview?a=v&q=cache:dhkeO1cTaMQJ:www.staff.unibe.ch/airoidi/stat_ind.pdf+SPSS+%22comparison+of+slopes+%22+filetype:pdf&hl=ja : 代表的市販統計ソフトパッケージ15種類の収載検定比較

Web04) <http://www.hku.hk/ecology/staffhp/kl/Biom12.ppt>: Mathematical description of the relationship between two variables using regression - Lecture 13 BIOL2608 Biometrics

Web05) <http://stateasy.com/>: StatEasy

Web06) <http://courses.washington.edu/qsci483/Lectures/30.pdf> : Lecture 30: Qualitative Predictor Variables

文 献

1) 瀧口 徹：歯科疫学統計 ー第4報 傾向(トレンド)分析の基礎ー, Health Science and Health Care,

- No 1 vol 6 : 53-67, 2006.
- 2) 瀧口 徹：歯科疫学統計 - 第5報 傾向（トレンド）分析の基礎 その2, Health Science and Health Care, No 1 vol 8 : 19-33, 2008
 - 3) 高德幸男, 瀧口 徹, 小林清吾, 矢野正敏, 筒井昭仁, 張 旌旗, 関原 敬, 堀井欣一：咀嚼機能に及ぼす加齢の影響について, 日本咀嚼学会雑誌, vol 4 : 41-50, 1995.
 - 4) 瀧口 徹：歯科疫学統計 - 第3報 重回帰分析, 多重ロジスティック回帰分析モデルの適合度判定指標の解釈 - - SPSS, STATAの利用に際して -, Health Science and Health Care, No 1 vol 5 : 35-49, 2005.
 - 5) 瀧口 徹：歯科疫学統計 - 第2報 一般化線形モデルの意義と潮流 -, Health Science and Health Care, No 1 vol 4 : 29-36, 2004.
 - 6) 赤池弘次：AICとMDLとBIC、オペレーションズ・リサーチ：経営の科学, 41 (7), 375-378, 1996.
 - 7) 丹後俊郎, 山岡和枝・高木晴良：ロジスティック回帰分析 SASを利用した統計解析の実際, 朝倉書店, 東京都, 第6刷, 2002. p200-201頁.
 - 8) Joël Ladmxarb Kati Castetbon, brc Valériane Leroy, et al: Pregnancy, body weight and human immunodeficiency virus infection in African women: a prospective cohort study in Kigali (Rwanda), Int. J. Epidemiol., 27, 1072-1077, 1998.
 - 9) 春日邦子, 十川知子, 北山秋雄, 時技正吉, 市村博, 太田原美作雄：平行線定量法を用いた風疹のEJISA法による抗体測定, 千葉衛研報告 第12号 37-40, 1988.
 - 10) A Abbott, M.L. Hayes: The Conditioning Role of Saliva in Streptococcal attachment to Hydroxyapatite Surface, J. Gen. Microbiol., 30, 809-816, 1984.
 - 11) Michael J. Crawley (野間口謙太郎 他訳)：統計学：Rを用いた入門書 (Statistics : An Introduction using R), 共立出版, 東京, 第1刷, p1-344, p111-165, 2008.
 - 12) 小野寺孝義, 山本嘉一郎：SPSS-Output-Pedia: Base SPSS辞典 (Base編) ナカニシヤ出版, 京都市, 第1刷, 145-163, 2004.
 - 13) John M. Last編, 重松逸造, 春日 齊, 柳川 洋監訳：疫学事典, 日本公衆衛生協会, 東京, 第1刷, p79-80, 1987.

A review of oral epidemiological statistics

ー Part VI : The bases of Trend Analysis 3rd follow-up review ー

Toru Takiguchi

(Fukai Institute of Health Science)

Key words : Trend Analysis, Linear & Non-linear Regression Line, Comparing of Slopes and Intercepts of Two Regression Lines, Free Statistical Software, Statistical R Software Environment

Abstract: The two best statistical software systems freely available under General Public License are R and Maxima. In this review the significance and usefulness of these two free software environments were highlighted through trend analyses of dental epidemiological data in the People' s Republic of China, thereby reinforcing and increasing understanding of trend analyses, which were thoroughly dealt with in the fourth and fifth reviews.

Contents of this review:

- (1) Significance and advantages of free statistical software such as R and Maxima.
- (2) Identity tests for slopes and intercepts of two regression lines.
- (3) Calculation of coefficients for all kinds of regression models.
- (4) Selection of optimal "goodness of fit" model based on biological plausibility.
- (5) Point estimation (e.g. number of present teeth of 100-year-olds) using an interpolation/extrapolation technique in R and Maxima.

Health Science and Health Care 9 (1) : 4 – 18, 2009