ベースラインを持つバランス型成長曲線モデルの推測について

佐藤健一[1]　冨田哲治[2]
[1]広島大学・原爆放射線医科学研究所　[2]県立広島大学・経営情報学部

すべての個体の測定時点が揃っているバランス型の経時測定データの例として Potthoff and Roy (1964) によって提示された歯科矯正データを表１に示す．このデータは統計ソフト R の nlme パッケージの中に Orthodont データとして格納されており，藤井ら (2015) において交互作用項の解析例として利用されている．

表１のデータにおいて女性をコントロール群として，男性の効果，つまり，男性の女性からの差分である性差に関心があるとしよう．このとき，女性の成長曲線の特定には大きな関心はなく，男性の成長曲線との差およびその有意性，あるいはその信頼区間には関心がある．このような状況は生存時間解析において広く利用されているコックスの比例ハザードモデルでも想定されており，ベースラインの生存曲線を特定することなく，共変量の効果の統計的推測が行われている．本発表では，Satoh and Tonda (2016) によって提案されたバランス型の経時測定データにおいてベースラインの成長曲線を特定することなく，共変量の効果を推定する方法を表１のデータに適用する形で紹介する．

今，個体数を$n = 11 + 16 = 27$，測定時点数を$p = 4$，Yを表の 3 列目から 6 列目のデータで作る$n \times p$観測値行列，ベースラインを少女のグループだけ ($r = 1$) とし，$U = \mathbf{1}_n$を長さn成分が 1 の$n \times r$行列，測定時点ごとのベースラインのパラメータを$1 \times p$パラメータ行列Λ，表１の 2 列目のデータ ($k = 1$) で作る$n \times k$行列をA，とすれば，ベースラインを持つ成長曲線モデルは，

$$\underset{n\times p}{Y} = U \underset{r\times p}{\Lambda} + A \underset{k\times q}{\Theta} X + \mathcal{E}, \qquad \mathcal{E} \sim N_{n\times p}(O_{n\times p}, \Sigma \otimes I_n),$$

とかける．ここで，行列Xは性差に対して直線を仮定すると，

$$X = \begin{pmatrix} 1 & 1 & 1 & 1 \\ 8 & 10 & 12 & 14 \end{pmatrix},$$

とかける．未知パラメータ行列はΛ，Θおよび測定時点間の共分散行列Σである．Σについては簡単のため既知とするが，未知であっても推定方法が提案できる．ベースラインについては大きな関心がないため，Satoh and Tonda (2016) はΛを推定することなく性差に関するΘを推定する方法として次の推定量を提案している．

$$\hat{\Theta} = \{A'(I_n - P_U)A\}^{-1}A'(I_n - P_U)Y\Sigma^{-1}X'(X\Sigma^{-1}X')^{-1}.$$

一方，Θが推定されれば，これを用いてベースラインの推定量も次のように与えることができる.

$$\hat{\Lambda} = (U'U)^{-1}U'(Y - A\hat{\Theta}X).$$

表１．少女 11 人，少年 16 人の 4 時点（8，10，12，14 歳）における歯科矯正データ．歯科矯正の診断に利用するために頭部 X 線写真から２つの基準点（脳下垂体と翼上顎裂）の距離(mm)が各年齢ごとに計測されており，性別の列は男性なら 1，女性なら 0 と示されている（出典，Potthoff and Roy (1964)）

番号	性別	8 歳	10 歳	12 歳	14 歳
1	0	21	20	21.5	23
2	0	21	21.5	24	25.5
3	0	20.5	24	24.5	26
4	0	23.5	24.5	25	26.5
5	0	21.5	23	22.5	23.5
6	0	20	21	21	22.5
7	0	21.5	22.5	23	25
8	0	23	23	23.5	24
9	0	20	21	22	21.5
10	0	16.5	19	19	19.5
11	0	24.5	25	28	28
12	1	26	25	29	31
13	1	21.5	22.5	23	26.5
14	1	23	22.5	24	27.5
15	1	25.5	27.5	26.5	27
16	1	20	23.5	22.5	26
17	1	24.5	25.5	27	28.5
18	1	22	22	24.5	26.5
19	1	24	21.5	24.5	25.5
20	1	23	20.5	31	26
21	1	27.5	28	31	31.5
22	1	23	23	23.5	25
23	1	21.5	23.5	24	28
24	1	17	24.5	26	29.5
25	1	22.5	25.5	25.5	26
26	1	23	24.5	26	30
27	1	22	21.5	23.5	25

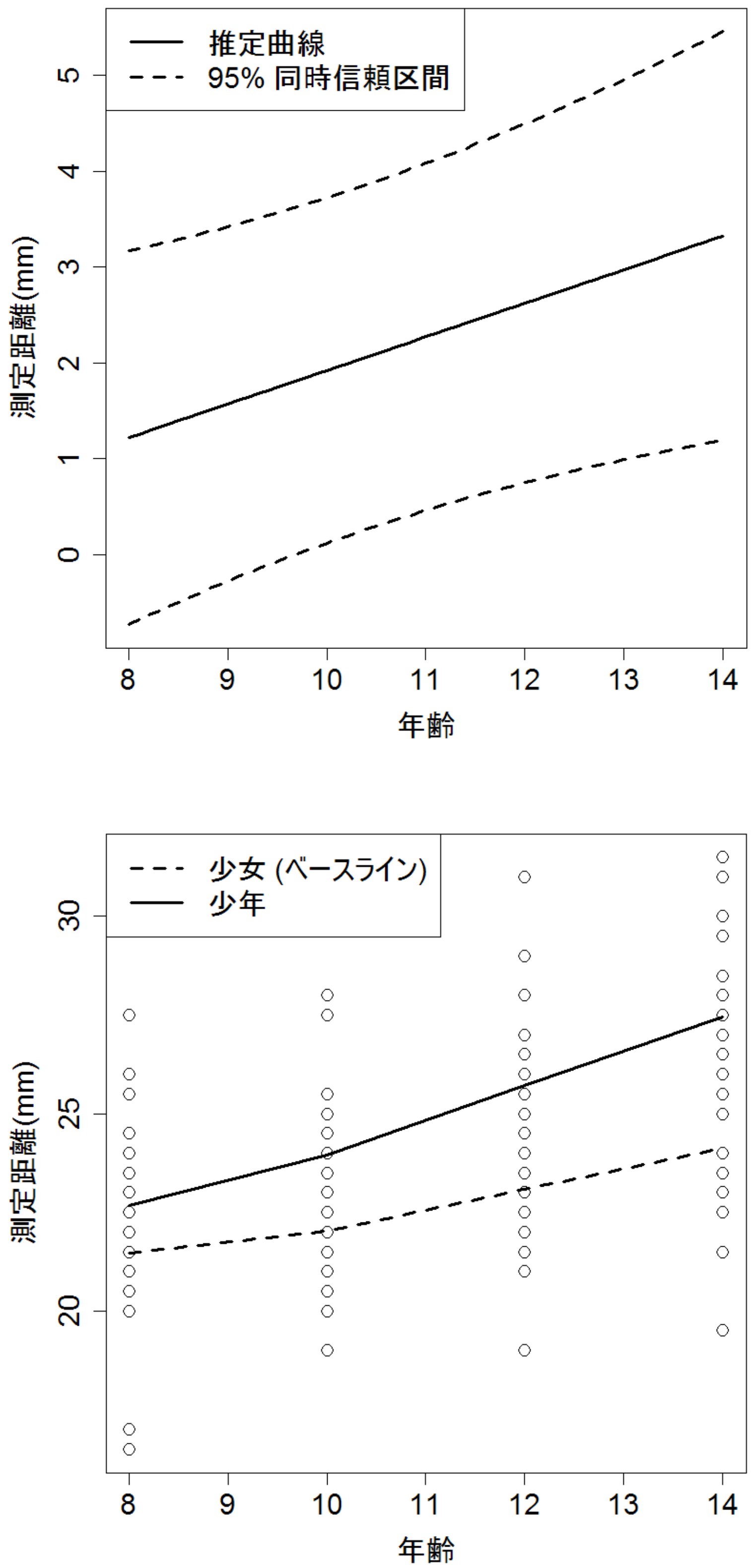

図１．（上）推定された性差と 95%同時信頼区間．（下）少女と少年の推定曲線．

提案手法によって推定された性差は図１（上）に，また，性差を推定した後で求められたベースライン，つまり，少女の成長曲線を図１（下）に示す．少女の成長曲線に性差を加えたものが少年であるので，これも図１（下）に付け加えた．図１（上）には Satoh and Yanagihara(2010)による線形な変化係数に対する同時信頼区間の導出方法を用いて性差の 95%同時信頼区間を与えた．

Potthoff and Roy(1964)によって提案されたモデルはベースラインのない場合（$U = O$）に対応している．実際，表１のデータでは測定時点数が$p = 4$と少ないため，少年少女，それぞれに直線を仮定しても，図１（下）とほぼ同じ推定曲線が得られる．一般的に，測定時点数が多くなってくると，ベースラインの曲線は複雑になりやすいが，共変量の効果については解析後に解釈する必要があるため自由度の高くない曲線が好まれる．本手法においては，「ベースラインには関心がない」という立場をとることで，ベースラインに対する曲線の基底の選択，あるいは，その適合度の吟味などから解放される．生存時間解析においてコックスの比例ハザードモデルが支持される理由の一つにベースラインの関数を特定する必要がないことが挙げられ，考慮すべきことが減るという点で解析者の負担が軽減されていると思われる．本手法にも同様の利点が期待される．

- 藤井良宣，佐藤健一，冨田哲治，和泉志津恵：医療系のための統計入門，事例でわかる統計シリーズ，実教出版，東京，2015.
- Potthoff, R. F. and Roy, S. N.: A generalized multivariate analysis of variance model useful especially for growth curve problems. *Biometrika*, 51, 313-326.
- Satoh K. and Tonda T.: Estimating regression coefficients for balanced growth curve model when time trend of baseline is not specified, *American Journal of Mathematical and Management Sciences*, in press.
- Satoh K. and Yanagihara H.: Estimation of varying coefficients for a growth curve model, *American Journal of Mathematical and Management Sciences* 30, 243-256, 2010.

連絡先：佐藤健一

ksatoh@hiroshima-u.ac.jp

http://home.hiroshima-u.ac.jp/ksatoh/

〒734-8551 広島市南区霞 1-2-3 総合研究棟 405 号室