近傍データに基づく児体重推定

久保武士[1)]　藤井恭一[1)]　赤塚孝雄[2)]　箕浦茂樹[3)]

1）筑波大学臨床医学系産婦人科　　2）筑波大学基礎医学系医工学　　3）東京大学産婦人科

「目的と意義」

子宮内の胎児の成熟度を正確に把握することは、周産期管理の上で、極めて重要である。児の未熟性は今なお、周産期死亡の大きな原因の一つであるばかりでなく、新生児期の呼吸異常や脳性小児まひ、あるいは未熟児網膜症の原因ともなることはよく知られている。

この未熟児は、早産や胎盤機能不全で生じるばかりではなく、時には妊婦管理の必要上やむをえず未熟児を娩出させなければならないこともしばしばある。

たとえば、心疾患を合併している妊婦を管理していて、妊娠の進行とともに心疾患が悪化し、母体の安全のために児の未熟性を承知で、予定日より早く児を娩出させなければならない時がそれである。母体の安全のためには、児を早く娩出させる必要があるが、しかしそれは、児にとって未熟なまま子宮外生活に適応しなければならないという不利を招く。児の未熟性を考慮し、妊娠を経続させることは、母体を危険にさらすというジレンマが生じる。近年胎児の成熟度を生化学的に、あるいは内分泌学的に判定する種々の方法が試みられているが、胎児の成熟度を示す端適な指標は、依然として胎児の体重である。従来産科医は、この児体重を推定するために、妊婦の腹部の触診所見や、子宮底長の計測値、あるいは腹部の計測値等を参考にしつつ主観と経験に基づいて直感的にこれを行なってきた。我々はこの主観と経験に基づくこの児体重推定を客観的で合理的なものにするために、子宮底長を始めとする種々臨床情報を説明変数とした重回帰式を作製し、これをマイクロコンピューターにプログラムして長年日常臨床に使用してきた。その結果は、必ずしも満足のゆくものではないが、人間が直感的に推定するのに比べればはるかに良い推定値が得られている。実際使用上の問題点の一つは、たとえば我々が一番正しい推定値を必要とするのは、児体重の極端に小さい未熟児や、逆に巨大児の場合であるが推定値の正確さという点では、両者共に平均体重近辺の予測に比べかなり推定の精度が悪くなるという事実である。一般に児体重が低いところでは、我々の予測式は児体重を高めに推定し、又極端に児体重が大きいところでは、むしろ児体重を低めに推定する傾向が認められている。このように平均値から著しくへだたったところでの推定精度が悪い原因の一つとして、重回帰式を作製するためのデータのサンプリングの問題が考えられる。平均児体重近辺のデータは多数集めやすいが、未熟児や巨大児に関しては、当然データが多数集まりにくい。従って平均値近辺では推定精度が高く、未熟児あるいは巨大児では推定精度が低くなるのは、ある意味では当然である。さりとて未熟児や巨大児のデータばかりを集めて重回帰式を作製するのもまた問題がある。児体重をたった一個の説明変数たとえば BPD (=Biparietal Diameter,児頭大横径) との関係を散

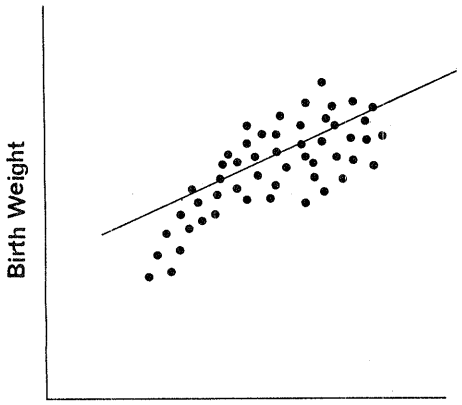


図1

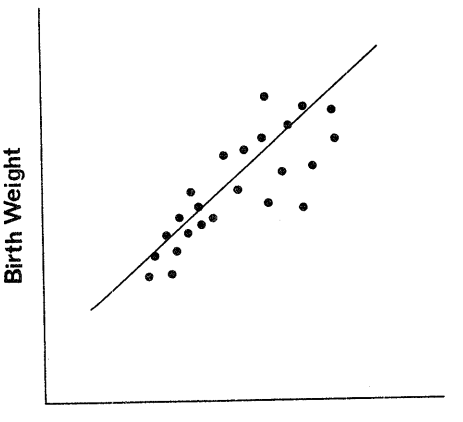
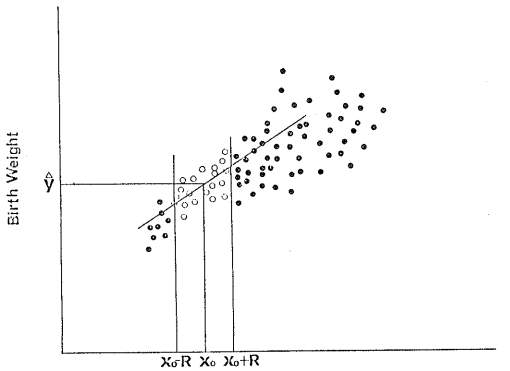


図2

布図に書いてみると、正常のデータ収集では図1のごとくなっている。従って、一本の直線でこの分布を近似すると当然BPDの低いところでは、児体重を高く推定しすぎることになる。これを意識的に児体重の低いものを中心に集めるとデータの分布は図2のごとくなり、それを近似する直線もまた図2のごとく最初のものとは異ったものになるであろう。この例に限らず、一般に医学のデータはそのサンプリングに問題が多い。本来は、実験計画法に基づき prospective にデータを集めるべきであろうが、これは言うべくして現実には実行できない事が多い。多くの統計学的分析の前提であるデータの無作為抽出は、医学においてはめったに満たされることのない条件である。一つ一つの症例が多くの労力と、長い時間をかけてやっと得られたという例が珍らしくない。さりとて無作為抽出でないデータをどのように細かく統計学的に分析しても、その分析結果の信頼度は低いものになるであろうし、ましてやその分析結果を他のデータに布衍することは、一層大きな誤診を招く危険がないとはいえない。このような難点を解決するために、我々は予測する点の近傍におけるデータだけを使ってその点における児体重の推定するという試みを行った。一変数の場合に限って説明すると、図3のごとく、点 X_0 における児体重を推定するには、の近傍すなわち X_0-R と X_0+R の間に入るデータだけを集め、そのデータだけに基づく回帰直線をひき、それによって X_0 における児体重を求めようというわけである。一個の重回帰式で全てを予測するかわりに、予測する点ごとの回帰式を求め、それによって推定を行なおうというアイデアである。



Estimation of Birth Weight by Regression Line derived from Date near the Point X_0

図3

「データ」

筑波大学産婦人科で得られたデータの完備した1443例をその分析の対象とした。データの完備したという意味は、個々の症例につき、目的変量である児体重を含めた以下の9変数につき完全な計測値が得られたものという意味である。説明変数とは、1）母体の年令　2）経産回数　3）母体身長　4）母体体重　5）分娩前二週間以内の子宮底長　6）分娩前二週間以内の腹囲　7）分娩前二週間以内に計測された胎児頭大横径すなわちBPD．8）胎児の在胎週数である。

「方法」

1．データの基準化

Y_i を症例 i の目的変量、すなわちこの場合は児体重、$X_i^T=(X_{i1},\cdots X_{ip})$ を説明変量ベクトルとする。これらを次のように Normalize する。

Normalization of Original Data

i-th sample　　y_i,　　$x^t=(x_{i1}, x_{i2},\cdots\cdots,x_{ip})$

$i=1,2,\cdots\cdots,n$　n:sample size

p:number of variables

$$U_i=\frac{y_i-\bar{y}}{S_y}\qquad \bar{y}=\frac{1}{n}\sum_{i=1}^{n} y_i\qquad S_y=\left\{\frac{1}{n}\sum_{i=1}^{n}(y_i-\bar{y})^2\right\}^{\frac{1}{2}}$$

$$z_i^t=(z_{i1}, z_{i2},\cdots\cdots z_{ip})$$

$$Z_{ij}=\frac{x_{ij}-\bar{x}_{\cdot j}}{S_j}\qquad \bar{x}_{\cdot j}=\frac{1}{n}\sum_{i=1}^{n} x_{ij}\qquad S_j=\left\{\frac{1}{n}\sum_{i=1}^{n}(x_{ij}-\bar{x}_{\cdot j})^2\right\}^{\frac{1}{2}}$$

$j=1,2,\cdots\cdots\cdots,p$

2．説明変数ベクトル X_0 の近傍のデータの収集

今点 X_0 における Y を予測しようとしている時、まず X_0 を上の例に倣って基準化する。

Normalization of Estimation Point $X_0^t=(x_{01},\cdots,x_{0p})$

$$Z_0^t=(z_{01},z_{02},\cdots,z_{0p})$$

$$z_{0j}=\frac{x_{0j}-\bar{x}_{\cdot j}}{s_j} \qquad j=1,2,\cdots,p$$

3．続いて点 Z_0 の近傍のデータを次の式に従って求める。

Selection of Data near the Point $Z_0(=X_0)$

$$D_{Z_0}=\{Z_i \mid Z_i \in D,\ \|Z_i-Z_0\|<R\}$$

D: a set of original data $Z_i(=X_i)$

$$\|Z_i-Z_0\|=\{\sum_{j=1}^{p}(z_{ij}-z_{0j})^2\}^{\frac{1}{2}}$$

すなわち点 Z_0 を中心とし半径Rの球内にある Z_i を集める。

4．この Z_i の集合 D_{Z_0} に基づいて、重回帰式を以下のように求める。

Beta Coefficients of the Multiple Regression Equation

f_{Z_0} derived from D_{Z_0}

$$f_{Z_0}(Z)=\hat{\beta}_1 z_1+\hat{\beta}_2 z_2+\cdots+\hat{\beta}_p z_p$$

$$(\hat{\beta}_1,\hat{\beta}_2,\cdots,\hat{\beta}_p)^t=(z^tz)^{-1}z^tU$$

$$z^t=(Z_{i1}^t,Z_{i2}^t,\cdots\cdots,Z_{ik}^t)$$

$$Z_{i_l}\in D_{Z_0} \qquad l=1,2,\cdots,k$$

k:number of samples which are included in D_{Z_0}

$$U^t=(U_{i1},U_{i2},\cdots,U_{ik})$$

U_{i_l}: U_i which corresponds to Z_{i_l}

5．これから最終的な目的変量 $\hat{Y}$ は次式に従って計算される。

Estimation of y at the Point X

$$\hat{U}=f_{Z_0}(Z_0)$$

$$\hat{y}=\bar{y}+S_y\hat{U}=\bar{y}+S_y\cdot Z_0'(z^tz)^{-1}z^tU$$

* Mean Square Error $=\sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i-\hat{y}_i)^2}$

今回は X_0 としてすべての X_i を選びすべての予測を行なった。すなわち internal check だけを行なった。

「結果及び考察」

表1に計算結果を示す、横軸は半径 Rであり、縦軸はその半径で推定した時の重相関係数その時の Mean Square Error, その半径で計算可能だったsampleの個数及び各点におけるその半径の球内に入ってきたsample数の平均値 $\bar{K}$ である。重相関係数は半径Rが50から減少するに従って、しばらく0.75を保ち半径Rが35の点でやや減少するがそれから再び漸増して、半径6では0.82にまで増加する。Mean Square Error は半径50で338であるが、半径が小さくなるに従ってしだいに小さくなり、半径6では243迄減少する。

最初の値の約７２％である。もっとも半径６で実際に推定されたsampleの個数は１０９個で、残り３６個の点では、sample 数が不足のために推定されていない。半径を５０にとるとすべての点において推定が可能であるが、しかし個々の点を推定するのに必要なsampleの平均数は１４０個で、データの中にはずい分隔たった点群が存在することがわかる。個々の点の半径Rの球内に入ってくる標本数の平均値は、当然半径が短かくなるに従って減少するが、半径６では平均２４個である。重回帰分析を行なうには、常識的に、少なくとも説明変数の３倍ほどのデータが必要であるといわれているが、これからすると半径６あるいは７のあたりが８個の説明変数で重回帰式をつくるときのRの下限だと思われる。以上のように半径を短かくとり、推定点の近傍のデータだけを使って推定することにすれば、実測値と推定値の相関係数は高くなり、Mean Square Error も少なくなって、推定精度の高くなる事が予想される。

冒頭にも述べたように、医学データの統計解析では sample size の小さいことが多い。児体重を重回帰式で推定する我々の問題にしても同様で、この点に対処するためにデータを集積しながら重回帰式を修正することは可能である。On-line端末からXを入力して予測値$\hat{Y}$を得るだけでなく、実測値Yが判明した時点で再びこの値を入力すると、この新しいデータを組み込んだ新しい重回帰式が計算され、次回の使用を待つ――というシステムは簡単につくれる。ところが、このようにして得られる重回帰式は、データの収集の仕方―― sampling に強く依存している。データを増やせば$\hat{\beta}$が母係数βに収束するという保証は少ない。しかし今回の結果によると、データを多数集積するに従ってデータの密度は高くなり、どの点をとってもその近傍には推定に必要なデータが充分集まって、精度の高い推定が可能になると思われる。尚、データの近傍をとる時の距離の問題は、このようなユークリッド距離がはたして妥当かどうかは問題であるが、今回はこの方法により検討した。

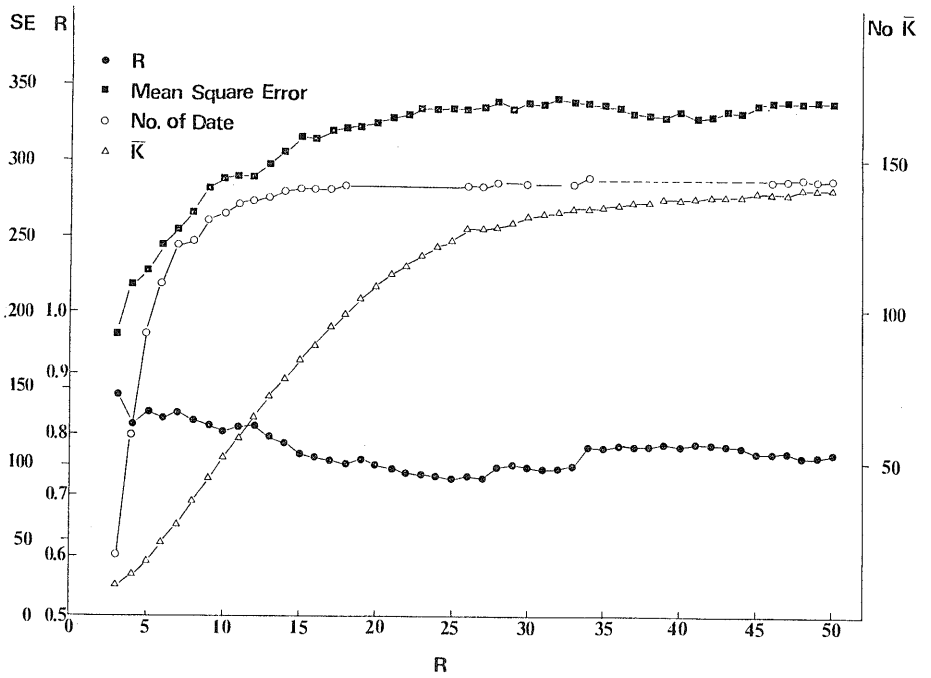


図４