実践講座

予測手法

(3)：統計的方法

上田　徹

3．統計的方法

ここでは統計的方法として回帰分析と数量化理論I類，II類をとりあげる．

3．1　回帰分析

3.1.1 回帰分析の概要

世帯数が多い地域では，電話加入者数が多いという傾向がある．この傾向は，世帯数をx，電話加入者数をyとするとxに関する単調増加関数$f(x)$を使って

$$y=f(x)+\varepsilon\ ;\ \varepsilon：誤差$$

で一般的に表現できる．特にyとxの間に直線的な関係を想定できるならば，

$$y=a+bx+\varepsilon \tag{3.1}$$

と表わせる．εが小さいほどyとxの間の関係はうまく表現できていることになる．地域ごとに世帯数のデータと電話加入者数のデータが得られているときは，それらを使ってεがなるべく小さくなるようにaとbが決められる．具体的には，地域i (i=1, 2, ..., n)の世帯数をx_i，電話加入者数をy_iとすると自乗誤差和

$$S_e=\sum_{i=1}^{n}\varepsilon_i^2=\sum_{i=1}^{n}\{y_i-(a+bx_i)\}^2 \tag{3.2}$$

が最小となるようにaとbが決められる．このように説明変数xを使って被説明変数yを表現するとき

「yのxへの回帰」

と言う．説明変数は独立変数，被説明変数は基準変数，従属変数とも呼ばれる．またa，bは偏回帰係数 (partial regression coefficient)と呼ばれるが，偏という言葉は誤解を生みやすく[1]，単に回帰係数と呼ばれる場合も多い．

このように，yを説明する変数が1つだけの場合を特に「単回帰」といって，説明変数が2つ以上の場合の「重回帰分析」と区別する場合もある．

うえだ　とおる　NTT通信網総合研究所
〒180 武蔵野市緑町3-9-11

3.1.2 回帰係数の推定法

(1)　無条件の場合

p個の説明変数$X_1, X_2, \ldots, X_p$の時点（あるいは地域など）iでの値を$x_{1i}, x_{2i}, \ldots, x_{pi}$とし，被説明変数$Y$の時点$i$での値を$y_i$ ($i=1, 2, \ldots, n$)とし，

$$y_i=a_0+a_1x_{1i}+a_2x_{2i}+\ldots+a_px_{pi}+\varepsilon_i \tag{3.3}$$

における回帰係数$a=(a_0, a_1, a_2, \ldots, a_p)'$を

$$S_e=\sum_{i=1}^{n}\varepsilon_i^2$$
$$=\sum_{i=1}^{n}\{y_i-(a_0+a_1x_{1i}+a_2x_{2i}+\ldots+a_px_{pi})\}^2 \tag{3.4}$$

を最小にするように決める．それには式(3.4)をa_iで偏微分した式を零とした$(p+1)$個の連立方程式（正規方程式という）を解けばよい．式(3.3)をベクトル，行列表現すると

$$y=Xa+\varepsilon \tag{3.5}$$

ただし，

$$y'=(y_1, y_2, \ldots, y_n),\quad \varepsilon'=(\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_n) \tag{3.6}$$

$$X=\begin{bmatrix} 1 & x_{11} & x_{21}\cdots x_{p1} \\ 1 & x_{12} & x_{22}\cdots x_{p2} \\ \multicolumn{3}{c}{\cdots\cdots\cdots\cdots\cdots} \\ 1 & x_{1n} & x_{2n}\cdots x_{pn} \end{bmatrix} \tag{3.7}$$

であり，正規方程式は

$$X'Xa = X'y \tag{3.8}$$

である．これは逆行列$(X'X)^{-1}$を使って

$$a=(X'X)^{-1}X'y \tag{3.9}$$

として求めることができるが，式(3.8)は線形連立方程式であり，中学校で習う連立方程式の解法（消去法）を用いた方がプログラムが簡単である．なお，Xの第1列を除去すれば定数項a_0のないモデルとなる．y_iの推定値$\hat{y}_i$はXaのi番目の要素$(Xa)_i$として与えられる．

説明変数間に高い相関性があるときには$|X'X|$が零に近くなり，回帰係数の推定は非常に不安定となる．

この推定困難な状況は多重共線性と呼ばれる．

（例3.1） x,y,z の間に

$$y=ax+\varepsilon,\ x=bz$$

の関係があるときに，yをx,zに回帰させようとすると任意のcに対して

$$y=(ac)x+\{ab(1-c)\}z+\varepsilon$$

が成り立ち，x,zの回帰係数は一意に決定できない．

このように多重共線性（変数間相関）があるときは

- 相関の強い説明変数の一部を削除してみる
- $(X'X)$のかわりに$(X'X+kI)$を用いる．kの値をいろいろふらしてみて回帰係数の推定値がある程度安定したら，それを採用する（この方法はリッジ[Ridge] 回帰と呼ばれる）

などの手段を講じなければならない．

また，(i)誤差εの平均が零（不偏性），(ii)分散が一定（等分散）かつ (iii)系列的にε_iとε_jは独立な (iv)正規分布に従うときに，式（3.9）で推定される係数aは分散が最小かつ不偏であり，そのときaは最良線形不偏推定量(BLUE)と呼ばれる．これらの仮定が満たされているか吟味しろという人もいるが，データにフィットする直線を求めるのだという発想に立てばあまり神経質になる必要はないと思う．ただし，その場合には各種検定を行なう際の前提も怪しいということを認識しておく必要がある．もしεの分散構造がはっきりわかっている場合には，一般化最小自乗推定量[2]を用いることができるが，そのようなことはまれであろう．

（2）条件つきの場合

回帰係数の符号が定性的に正か負か（説明変数に関して単調増加かどうか）がわかっている場合に，式（3.9）で得られる係数の符号が逆転しているときにはどうすればよいだろうか？

説明変数間に相関が強い場合には符号が逆転する説明変数がでることがある．たとえば例3.1で $\{a,\ b>0\}$ のときyはxに関してもzに関しても単調増加である．しかし，$\{c>1\}$とすればzの係数は負となってしまい，zに関する単調増加性を実現しないことになってしまう．

条件を満たさない説明変数をすべて削除し，残された変数だけで回帰分析し，その結果，得られた回帰係数が条件を満足していればそれで良しとすることが考えられる．しかし削除後の回帰係数が条件を満たす保証はないし，削除するのは一部でいいかもしれない．逐次，説明変数を削除，追加し，よさそうな変数の組合せを考えたらと思うが，変数の数が多くなるにつれて現実的な対応ではなくなる．非負条件だけでなく，より一般的な係数に関する線形不等式制約（ただし等号を含む）のある場合にも文献[3]には述べられているが，ここでは比較的利用度の高いと思われる係数非負条件を考慮した最小自乗法（以下では"NNLS"と略す）を紹介する．ただし，非負条件は説明変数x_iを$(-x_i)$とすることにより非正条件のある係数にも適用できる．

[NNLSアルゴリズム]

①非負条件のない回帰係数番号の集合をQとし，回帰係数非負条件を満たす回帰係数番号の集合をP，満たさない回帰係数番号の集合をZ，回帰係数の集合を$\boldsymbol{a}$と定義する．

②初期値としてPを空集合，ZをQにない回帰係数番号の集合，$\boldsymbol{a}'=(0, 0, ..., 0)$ とする．全変数を使って回帰分析をし，その結果から回帰係数非負条件を満たす回帰係数番号をPの要素とする．得られた回帰係数で負のものは零，そうでないものはそのままとした回帰係数の集合を$\boldsymbol{a}$の初期値とすることも考えられる．

③ $w_j \triangleq \partial S_e/\partial a_j$

$$=-\sum_{i=1}^{n} 2x_{ji}\{y_i-(a_0+a_1x_{1i}+a_2x_{2i}+\ldots+a_px_{pi})\}$$

$$(j=0, 1, 2, ..., p)$$

を計算する．[w_jは回帰係数a_jに関する自乗誤差和の傾向（増加か減少か）を表わす]

④Zが空集合またはすべてのj $(\in Z)$ に対して$w_j>0$ (a_jを0から正値に増やしてもS_eは減らない）ならばそれまでに求めた$\boldsymbol{a}$を最少自乗解として終了．

⑤$w_t=\min\{w_j : j\in Z\}$となるような回帰係数番号$t(\in Z)$ を求める．（S_eの減少分が最も大きい回帰係数番号tを求める）

⑥ tをZからPに移す

⑦PおよびQ内の要素に対応する変数だけを使って最少自乗法を適用し，得られた回帰係数とZ内の要素に対応する零の値を持つ回帰係数からなるベクトル$\boldsymbol{b}=(b_0, b_1, b_2, ..., b_p)$ を求める．

⑧すべてのj $(\in P)$に対して$b_j>0$ ならば$\boldsymbol{a}$を$\boldsymbol{b}$と置換して③に戻る．

⑨ $\alpha \triangleq a_q/(a_q-b_q)=\min\ \{a_j/(a_j-b_j) : b_j<0,\ j\in P\}$ となるq $(\in P)$, αを求める．

⑩ $\{a_j-\alpha(a_j-b_j)\}$ を改めてa_jとする．[$a_q=0,\ a_j\geq 0,$

$(j \in P)$である]
⑪ $a_j=0$である$j\,(\in P)$をPからZに移し，⑦へ戻る．

ステップ④で得られた$\boldsymbol{a}$は

$$\{a_j>0 : j \in P\} \quad \text{および} \quad \{a_j=0 : j \in Z\}$$

を満足する．

3.1.3 変数，モデルの選択

変数，モデルの選択を機械的に行ないやすいのは赤池の情報量規準AICであろう．AICは

$$\begin{aligned} \mathrm{AIC} = & -2\times(\text{モデルの最大対数尤度}) \\ & +2\times(\text{モデルの自由パラメタ数}) \end{aligned} \tag{3.10}$$

で定義され，AICが小さいほどよいモデルとされる．右辺第1項がモデルのあてはまりの悪さ，第2項が説明変数の数についてのペナルティに相当し，モデルの適合度が同程度ならば変数は少ないほどよいという「ケチの原理 (principle of parsimony)」にもとづく考え方の具体化といえる．

式(3.3)におけるε_iがiにかかわらず同じ正規分布に従うときには

$$\mathrm{AIC}=n(\log_e 2\pi+1)+n\log_e S_e/n+2(p+2) \tag{3.11}$$

となり，右辺第2項がS_eに関して単調増加であることからあてはまりの悪さを表現していることがわかるであろう．また，AICは真のモデルからの距離を計る量であり，真のモデルがわからないことから間隔尺度（2つのAICの間の差分）にのみ意味がある．

式(3.11)には正規分布の前提が必要であったが，他のほとんどの検定方法も正規分布の前提を必要とすること，式(3.11)右辺第2項が最小自乗法と適合しやすい量であることから相対的比較という意味では正規分布の前提に神経質になる必要はない．ここで，AICは過剰適合 (overfitting) モデルを選択しやすい（パラメタ数が大目にでる）という指摘があるが，それに対しては式(3.10)の第2項の係数2の代わりに$\log n$ (n：データ数) を用いる尺度も提案されている．要するに，AICの小さな差で一方のモデルを選ぶことには危険があり，AICの小さいものから並べられたいくつかのモデルから定性的に望ましいモデルを選ぶ必要がある．

変数，モデルの選択には分散分析や偏相関係数，標準回帰係数を用いる方法があるが，多くの成書で述べられているのでここでは述べない．

3.1.4 通信分野における応用例

ここでは通信分野における3応用例を示し，読者の参考としたい．

（1）発生呼数の直交多項式回帰

電子交換機における3分間ごとの発生呼数データyに対して曲線回帰を行なってみる．p次多項式をあてはめて

$$\hat{y}=b_0+b_1x+b_2x^2+\ldots+b_px^p \tag{3.12}$$

とし，p個の説明変数x^i $(i=1,2,..,p)$を用いた重回帰分析を行なうことが考えられるが，その場合には次数を$(p+1)$次としたとき，改めて偏回帰係数b_iを推定し直さなければならない．しかし，直交多項式を用いると低次の多項式の推定結果を高次多項式の推定に利用することができるので，ここではその概要を示すとともに分析例を示す．

時点t $(t=1,2,..,n)$におけるi次の多項式X_iをX_{it}と表わしたとき

$$\sum_{t=1}^{n} X_{it}X_{jt}=0 \quad (i \neq j) \tag{3.13}$$

を満たすならば各多項式は互いに直交していると言う．次のような逐次的な方法でX_{it}の値を決めることができる[4]．すなわち

$$X_{it}=(x_t-\beta_t)X_{i-1,t}-\gamma_t X_{i-2,t}$$
$$\beta_t=\sum_{t=1}^{n} x_t X_{i-1,t}^{\ 2} \Big/ \sum_{t=1}^{n} X_{i-1,t}^{\ 2}$$
$$\gamma_t=\sum_{t=1}^{n} X_{i-1,t}^{\ 2} \Big/ \sum_{t=1}^{n} X_{i-2,t}^{\ 2} \qquad (i=2,3,\ldots.)$$
$$X_{0t}=1,\quad X_{1t}=x_t-\bar{x},\quad x_t=t,\quad \bar{x}=n(n+1)/2$$

とすればよい．このときy_iの推定値$\hat{y}_i$は

$$\hat{y}_i=\sum_{j=1}^{p} C_jX_{ji} \tag{3.14}$$

$$C_j=\sum_{i=1}^{n} y_iX_{ji} \Big/ \sum_{i=1}^{n} X_{ji}^{\ 2} \tag{3.15}$$

で与えられる．すなわちC_jはX_{ki} $(k \neq j)$には依存していない量である．

表3.1のデータに対して式(3.11)によりAICを求めると**表3.2**のようになり，AIC最小化基準で次数pを決定すると4次が最適となる．

（2）AT&Tにおけるサービス別需要予測[5]

市内サービス，市外サービス，発信無料サービス，専用線サービス別に次式により需要を予測する：

$$\begin{aligned} \log Q_{it} = & a_{i0}+a_{i1}\log Q_{i,t-1}+a_{i3}\log(Y_{it}/P_t^*) \\ & +a_{i4}\log(P_{it}/P_t^*)+a_{i5}\log(Z_{it}) \end{aligned}$$

$Q_{it}=R_{it}/P_{it}$：サービスiの時点tでの需要指標
R_{it}：サービスiからの時点tでの収入
P_{it}：サービスiの時点tでの価格指数

表 3.1 観測値

t	y_t	t	y_t
1	1934	11	2026
2	2066	12	1954
3	2174	13	2082
4	2198	14	1928
5	2218	15	2011
6	2190	16	1874
7	2130	17	1849
8	2146	18	1871
9	2249	19	1926
10	1994	20	1794

表 3.2 AIC

p	AIC
0	256.8
1	244.6
2	238.9
3	230.8
4	228.6
5	230.4

$Q_{i,\,t-1}$：サービス i の時点 $(t-1)$ での需要指標であるが，習慣形成の指標として用いられている．

Y_{it}：可処分所得

P_t^*：一般価格デフレータ（割引率）

Z_{it}：市場変数

説明変数として右辺に現われている$Q_{i,\,t-1}$はサービス i の時点 $(t-1)$ での需要指標であるが，非耐久財における習慣形成を表わしており，現在の需要は習慣の心理的ストックからプラスの影響を受けていることを意味する[6]．このようなラグ付き変数の使用は文献[6]に多くの例が紹介されている．

特に発信無料サービスでは，

$$\log Q_{Wt} = 1.92 + 0.83 \log Q_{W,t-1} - 0.14 \log (P_{it} / P_t^*) + 0.18 \log (Z_{Wt})$$

Z_{Wt}：国内航空会社販売額

となっており，Z_{Wt}はこのサービスが主に航空機予約に使われていることを示している．

（3）カナダ発信の国際電話トラヒック予測[7]

時点 t での呼数M_t は次式によって推定される．

$$\log M_t = -a + (1-b) \log M_{t-1} + c \log (Y_t + Y_{t-1}) - d \log P_{in,t} + e \log EXR_t + f \log IMP_t + \Sigma_i\ g_i\ \delta_{it}$$

Y_t：時点 t でのGNP

$P_{in,t}$：時点 t での最初の３分間の料金

EXR_t：時点 t でのカナダ・ドルとポンドの交換レイト

IMP_t：時点 t での商品輸入量

δ_{it}：ストライキや季節に対応するダミー変数

このモデルが選ばれるまでには，他の説明変数やラグの取り方などについて試行錯誤が行なわれており，文献[7]はモデル構成時の参考になる．

3．2 林の数量化理論

林の数量化理論のうち

I 類は質的変数を用いた回帰分析，

II 類は質的変数を用いた判別分析

で予測モデルと呼ばれ，

III類，IV類は対象を分類するための記述モデルで

多次元尺度構成法の１つ

ということもでき，ここでは前２者に限定する．また前２者は外的基準がある場合に相当し，外的基準が数量のときをI類，分類のときをII 類ということもできる．

3.2.1 数量化 I 類

数量化 I 類では，被説明変数Yの個体 i での値をy_iとすると，

$$y_i = b_0 + \Sigma_{j,k}\ b_{jk} \delta_{jk}(i) + \varepsilon_i \tag{3.16}$$

$\delta_{jk}(i)=1$：個体i が属性種別（アイテム）j の分類（カテゴリ）k に属するとき

0：その他

のように説明変数がダミー変数$\delta_{jk}(i)$で与えられる．

ダミー変数の特徴として，アイテムj ごとに１カテゴリK_j は他のどのカテゴリでもないということで特定化できるのでそのカテゴリに対応するパラメタは冗長であり，パラメタ b_{jK_j} を削除して冗長性を除去する．このようにして式 (3.9) の(X^tX) が正則化されるという点を除けばパラメタ推定は節 3.1.2 と同様に求められる．

このことから類推できるように説明変数には数値データ（量的変数）と分類データ（質的変数）とが混在してかまわない．混在時には数値の範囲を適当に区切って質的変数化するということもよく行なわれる．しかし，その場合には説明変数と被説明変数間の単調な関係がないかなどをチェックしなければならない．たとえば図 3.1 のように定量的にx_j として扱うときには傾きa_j が得られ，質的変数として扱うときにはカテゴリkごとにパラメタ b_{jk} が得られる．説明変数が広告量z_iで，被説明変数が販売額y_jのときにはz_iをそのままx_{ji}として用いるのではなく，{$f(z)$: zに関する単調増加関数}を用いた方がよい．そこで質的変数としてあつかったモデルから得られたb_{jk}からfの形を推定して量的変数 $f(z_i)$とし，改めて質的，量的変数混在下でパラメタ推定すべきである．

カテゴリ化で注意しなければならないのは，極度に

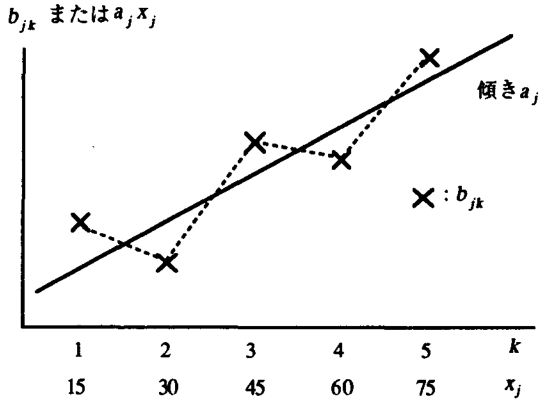

図 3.1 説明変数とパラメタ値

対応する個体数の少ないカテゴリを作らないようにしなければならないことである．アイテムJのカテゴリKを持つのは個体Iだけの場合には，最小自乗法によるパラメタ推定ではb_{JK}にかかわらず他のパラメタを推定でき，パラメタb_{JK}はε_Iを零にするように決めることができる．そのため，レンジ

$$R_J = \max_k b_{Jk} - \min_k b_{Jk} \qquad (3.17)$$

が極度に大きくなりうる．このようなことを避けるためにはカテゴリの統合も必要である．

変数，モデルの選択には，

$$v_{ji} = \Sigma_k b_{jk} \delta_{jk}(i)$$

を量的変数の場合の$\{a_j x_{ji}\}$の代わりとして考えれば節 3.1.3 と同様の議論が可能であるが，より簡単な方法として，レンジR_jの大きなアイテムjを有効なアイテムとする方法がよく用いられる．ただし，式 (3.17) の max，min を取るときには削除したパラメタの値が零であることも考慮しなければならない．

数量化I類は，たとえば世帯属性や事業所属性から通話量を推定するのに用いうる[8]．このとき，エリアA内の事業所についての式 (3.16) の和は

$$\Sigma_{i\in A} y_i = \Sigma_{i\in A}(b_0 + \varepsilon_i) + \Sigma_{j,k} b_{jk} \Sigma_{i\in A} \delta_{jk}(i)$$

となるので$\{\Sigma_{i\in A} \delta_{jk}(i)\}$はエリア$A$内で属性$\{jk\}$を持つ事業所数となり，エリアごとの統計情報を使ってエリアごとの通話量を推定・予測することができる．このような集計が可能なことは式 (3.16) の線形モデルの大きな利点である．

3.2.2 数量化II類

外的基準が分類で与えられている (y_iが直接観測できない) ことからレベル調整項b_0が必要でないことを除いては式 (3.16) と同じく，個体 i に対して数値

$$f_i = \Sigma_{j,k} b_{jk} \delta_{jk}(i) \qquad (3.18)$$

を与える．b_{jk}を求める基準として相関比

$$\eta^2 = \sigma_b^2/\sigma^2 = (fの層間分散)/(fの全分散) \qquad (3.19)$$

の最大化を考える．これは層（同一分類）内の分散をできるだけ小さく，層間分散をできるだけ大きくすることであり，層別（分類）を最も効果的に見せる数値f_iあるいはb_{jk}を求めることである．これは一般には固有方程式の 1 でない最大固有値η_1^2に対応する固有ベクトル

$$b_1{}^t = (b_{11}, b_{12}, \ldots, b_{21}, b_{22}, \ldots) \qquad (3.20)$$

を求めればよい．

特に{分類数H=2（2群)}のときには，連立方程式をbについて解けばよく，分類1，2に属する個体にそれぞれ値$(-n_2/N)$，(n_1/N)を与えてそれを被説明変数とした数量化I類による解と一致する（[9], p.96）．ただし，定数項の扱いや削除するカテゴリについてはI類とII類で同じモデルにしなければならない．固有方程式や連立方程式の具体的な形については[9]以外にも多数の成書があるので参照されたい．

2群のとき，個体（個人）i に対して式 (3.18) から推定されるf_i を用いて

・$f_i \geq f_0$ならば群2（たとえば「利用意向のある者」）
・$f_i < f_0$ならば群1（たとえば「利用意向のない者」）

と群分けする．ただし，群iに属する人数xの分布関数を$G_i(x)$とするとき，$\{G_1(f_0) = 1 - G_2(f_0)\}$となるように$f_0$は決められる．エリア$A$内の個人についての和は

$$\Sigma_{i\in A} f_i = \Sigma_{j,k} b_{jk} \Sigma_{i\in A} \delta_{jk}(i) \qquad (3.21)$$

となるので数量化I類の場合と同じくエリアごとの統計情報を使ってエリアごとの利用意向の強さを定量化でき，販売対象エリアの選別に利用することができる．なお，$G_1(f_0)$は境界f_0で群分けしたときの判別的中率になっている．

この利用例[10]としては，

(i)利用者($f_i \geq f_0$)，非利用者($f_i < f_0$)のそれぞれの属性がわかっているときに，$\{f_i \geq f_0\}$となっている非利用者iを販売対象とする，

(ii)自社製品と他社製品それぞれの利用者とその属性がわかっているときに，自社と他社の判別に寄与する属性から自社製品の長短を知り，新製品開発の参考とする

など，対象が分類されており，対象ごとの属性が判っ

ていれば利用できるものと，式(3.21)のようにマスに対する推定としての利用が図れるものとがある．

3．3　新サービスの予測法

新サービスの予測には既存サービスとの何らかの意味での類似性が用いられる．新サービスが既存サービスに現われている属性の新たな組合せであれば数量化I類を利用することが考えられる[11].

しかし，新しい機能の付与など既存サービスにない要因を持つ新サービスに対しては適用できなかったり，類似サービスが少なすぎるためにパラメタが特定サービスだけで決まったりすることがある（節3.2.1参照）．たとえば文献[12]では，

- 経済性要因としてイニシャルコスト，ランニングコスト，ソフトコスト，ブーム要因の4要因，
- 利便性要因として利便性，必需性，情報量，娯楽性の4要因

をとりあげ，サービスごとに各要因が望ましい水準に達しているか（1点），いないか（0点）を判断し，それと既存サービスの普及率とを関連づけて新サービスの普及率を推定するという手順を踏んでいるが，既存サービス数が少なすぎるため数量化I類が適用できず，もっと簡易な方法で推定している．

このように新サービスの予測は大雑把になりがちであるが，導入当初（期間 t_0）の需要データが利用できる場合には次のような方法が考えられる[13].

①既存サービス i は時点 T_i，新サービスは時点 T_N にサービス開始されたものとする．新サービスは期間 t_0 の間の需要データが利用でき，時点 $(T_N + t_0 + t_N)$ での予測をしたいものとする．

②サービス i の時点 t での需要モデルとして

$$y_i(t) = x_i(t)\,\boldsymbol{a}_i + \boldsymbol{\varepsilon}_i \quad (i = 1,2,\ldots,n-1,n,N)$$

を考え，期間 t_0 の間の需要データからパラメタ $\boldsymbol{a}_i$ の最小自乗推定量 $\overline{\boldsymbol{a}_i}$ を求める．ただし，t_0 は小さいので変数の数はごく少数に限られる．

③サービス $i\ (i = 1,2,..,n)$ の補正係数

$$c_i = y_i(T_i + t_0 + t_i) \,/\, \overline{y_i}(T_i + t_0 + t_i)$$

$$\overline{y_i}(T_i + t_0 + t_i) = x_i(T_i + t_0 + t_i)\,\boldsymbol{a}_i$$

を求める．

④新サービスと既存サービス i とのパラメタに関する類似度

$$D_i^{\,2} = (\overline{\boldsymbol{a}_i} - \boldsymbol{a}_i)'\,(\overline{\boldsymbol{a}_i} - \boldsymbol{a}_i)$$

を使って，新サービスの補正係数

$$c_N = (1 + \sum_{i=1}^{n} c_i / D_i^{\,2}) \,/\, (1 + \sum_{i=1}^{n} 1 / D_i^{\,2})$$

を求める．

⑤予測値

$$\overline{y_i}(T_N + t_0 + t_N) = c_N\, x_N(T_N + t_0 + t_N)\,\boldsymbol{a}_i$$

を求める．

参考文献

[1] M.G.Kendall and W.R.Buckland： " A Dictionary of Statistical Terms " , 4th edition ,1982, 千葉大学統計グループ訳「ケンドール統計用語辞典」，丸善, 1987.

[2] A.C. Aitken：" On Least Squares and Linear Combinations ", Proceedings of the Royal Society of Edinbiurgh , 55, pp.42–48, 1935.

[3] C.L.Lawson and R.J. Hanson ：" Solving Least Squares Problem " , Prentice-Hall, 1974.

[4] 伏見，赤井：「直行関数系」，節3.2，共立出版, 1981.

[5] B.E.Davis ほか："An Econome rtric Planning Model for American Telephone and Telegraph Company ", *The Bell Journal of Economics and Management Science*, 4, pp.29-56, 1973.

[6] H.S.Houthaker and L.D.Taylor ：" Consumer Demand in the United States, 1929–1970 " , Harverd Univ. Presss (1966), 黒田，西川，辻村 訳「消費需要の予測」，勁草書房, 1968.

[7] R.Khadem：" An Econometric Forecasting Model of the Demand for International Voice Telecommunication from Canada " , 8th International Teletraffic Congress, 212, 1976.

[8] W.J.Infosino："Relationships Between the Demand for Local Telephone Calls and Household Characteristics", *The Bell System Technical Journal*, 59, July-August, pp.931-953, 1980.

[9] 岩坪秀一：「数量化法の基礎」，朝倉書店, 1981.

[10]上田徹，藤原貢：「統計的手法にもとづく通信利用動向の分析」，*NTT R&D*, Vol.40, No.12, 1991.

[11]三上史明：「新商品販売予測における多変量解析の適用」，品質管理，Vol.35, No.5, pp.741-744, 1984.

[12]広帯域ISDNに関する調査研究会：「広帯域ISDNに関する調査研究　報告書」, 1992.

[13]H.Saito："A Demand Forecasting Method for New Telecommunication Services", *Journal of the Operations Research Society of Japan*, Vol.30, No.2, pp.248-262, 1987.