第11卷 第4期 湖北大学学报（自然科学版） Journal of Hubei University (Natural Science) Vol. 11 No. 4 .1989

中图分类号：P91

多变量对多变量线性回归方法及其在区域研究中的应用

刘妙龙

（区域规划研究所）

摘 要 本文探讨了区域研究中有广泛潜在用途的多变量对多变量线性回归分析方法，给出了模型，最小二乘解和计算实施过程；以湖北省人民生活水平与社会、经济诸条件为例，介绍了该方法的实际应用。

关键词 多变量，线性回归，区域研究。

众所周知，回归分析方法是区域研究中最常用，亦是最为有效的分析预测方法之一。但在区域研究（经济的或自然的）中，经常需要研究多个变量对多个变量之间的关系。这些变量，不仅在自变量与因变量之间，而且自变量本身，各个因变量之间均存在有某种程度的相关性。我们不仅需要研究某个特定的因变量与一组自变量之间的依赖关系，更需要探讨作为一个整体的一组（多个）因变量与一组自变量之间的制约关系。这只要在传统的多元线性回归方法基础上予以扩展，即通过对每个因变量与自变量之间的回归方程参数估计，联立求解，最终求出一组回归方程的参数估计值。其基本的求解方法仍然是最小二乘法。这种综合求解多个自变量与多个因变量之间线性相关关系的回归方法，称为多变量对多变量线性回归方法。

1 数学模型

有p个自变量x_1，x_2，……x_p，m个因变量y_1，y_2，……y_m的n次观测样本资料（$x_{\alpha 1}$，$x_{\alpha 2}$，……$x_{\alpha p}$，$y_{\alpha 1}$，$y_{\alpha 2}$，……$y_{\alpha m}$），$\alpha=1, 2, \cdots\cdots, n$，矩阵表达式为：

$$X_{(n\times p)}=\begin{pmatrix} x_{11} & x_{12}\cdots\cdots x_{1p} \\ x_{21} & x_{22}\cdots\cdots x_{2p} \\ \vdots & \\ x_{n1} & x_{n2}\cdots\cdots x_{np} \end{pmatrix}=(X_1, X_2, \cdots\cdots, X_p)=\begin{pmatrix} X'_{(1)} \\ X'_{(2)} \\ \vdots \\ X'_{(n)} \end{pmatrix}$$

本文1989年2月14日收到。

$$Y_{(n\times m)}=\begin{pmatrix} y_{11} & y_{12}\cdots\cdots y_{1m} \\ y_{21} & y_{22}\cdots\cdots y_{2m} \\ \vdots & \\ y_{n1} & y_{n2}\cdots\cdots y_{nm} \end{pmatrix}=(Y_1Y_2\cdots\cdots Y_m)=\begin{pmatrix} Y'_{(1)} \\ Y'_{(2)} \\ \vdots \\ Y'_{(n)} \end{pmatrix}$$

其中：

$$X_i=\begin{pmatrix} x_{1i} \\ x_{2i} \\ \vdots \\ x_{ni} \end{pmatrix} \qquad Y_j=\begin{pmatrix} y_{1j} \\ y_{2j} \\ \vdots \\ y_{nj} \end{pmatrix}$$

$$i=1,\ 2,\ \cdots,\ p, \qquad j=1,\ 2,\ \cdots,\ m$$

多变量对多变量线性回归的基本前提是因变量Y_α与自变量X_1，X_2，…，X_p之间存在有某种线性相关关系，可表达为：

$$Y_\alpha=\beta_{0\alpha}+\beta_{1\alpha}X_1+\beta_{2\alpha}X_2+\cdots+\beta_{p\alpha}X_p+\varepsilon_\alpha \qquad (1)$$

$$\alpha=1,\ 2,\ \cdots,\ m$$

即：

$$\begin{pmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_m \end{pmatrix}=\begin{pmatrix} \beta_{01} & \beta_{11} & \beta_{21}\cdots\beta_{p1} \\ \beta_{02} & \beta_{12} & \beta_{22}\cdots\beta_{p2} \\ \cdots & & \\ \beta_{0m} & \beta_{1m} & \beta_{2m}\cdots\beta_{pm} \end{pmatrix}\cdot\begin{pmatrix} 1 \\ X_1 \\ X_2 \\ \vdots \\ X_p \end{pmatrix}+\begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_m \end{pmatrix} \qquad (2)$$

简写为：

$$Y=(1X)\beta+\varepsilon \qquad (3)$$

式中：

$$(1X)=\begin{pmatrix} 1 & x_{11} & x_{12}\cdots x_{1p} \\ 1 & x_{21} & x_{22}\cdots x_{2p} \\ \vdots & & \\ 1 & x_{n1} & x_{n2}\cdots x_{np} \end{pmatrix} \qquad \beta=\begin{pmatrix} \beta_{01} & \beta_{11} & \beta_{21}\cdots\beta_{p1} \\ \beta_{02} & \beta_{12} & \beta_{22}\cdots\beta_{p2} \\ \cdots & & \\ \beta_{0m} & \beta_{1m} & \beta_{2m}\cdots\beta_{pm} \end{pmatrix}$$

β为回归系数估计矩阵。

ε为相应的随机残差矩阵，

$$\varepsilon=\begin{pmatrix} \varepsilon_{11} & \varepsilon_{12}\cdots\varepsilon_{1m} \\ \varepsilon_{21} & \varepsilon_{22}\cdots\varepsilon_{2m} \\ \cdots & \\ \varepsilon_{n1} & \varepsilon_{n2}\cdots\varepsilon_{nm} \end{pmatrix}=\begin{pmatrix} \varepsilon'_{(1)} \\ \varepsilon'_{(2)} \\ \vdots \\ \varepsilon'_{(n)} \end{pmatrix} \qquad (4)$$

其分量$\varepsilon'_{(\alpha)}=(\varepsilon_{\alpha1}\ \varepsilon_{\alpha2}\cdots\varepsilon_{\alpha m})$ $\alpha=1,\ 2,\ \cdots,\ n$相互独立，每个 $\varepsilon'_{(\alpha)}$的均值向量为0，协方差矩阵为V。显然ε随机，相应的Y的均值向量$\bar{Y}$也随机。所以多变量对多变量线性回归模型为：

$$\begin{cases} E(Y)=(1X)\beta \\ Y'_{(1)},\ Y'_{(2)},\ \cdots,\ Y'_{(n)}\text{互不相关，协方差矩阵}V>0 \end{cases} \qquad (5)$$

2 参数的最小二乘法估计

随机残差矩阵ε：

$$\varepsilon = Y - (1X)\beta$$

$$\varepsilon' \cdot \varepsilon = (Y - (1X)\beta)' \cdot (Y - (1X)\beta) \tag{6}$$

显然，在最小二乘原理下求解β的参数解β^*，矩阵$\varepsilon'\varepsilon$达到最小。求解β^*的正规方程组为：

$$\begin{pmatrix} 1' \\ X' \end{pmatrix} (1X)\beta = \begin{pmatrix} 1' \\ X' \end{pmatrix} \cdot Y \tag{7}$$

对矩阵$(1X)$，设其秩$r = p+1$，则

$$\begin{pmatrix} 1' \\ X' \end{pmatrix} \cdot (1X) = \begin{pmatrix} n & 1'X \\ X'1 & X'X \end{pmatrix} \tag{8}$$

逆阵存在，其逆可由分块矩阵求逆法求得为：

$$\begin{pmatrix} n & 1'X \\ X'1 & X'X \end{pmatrix}^{-1} = \begin{pmatrix} n^{-1} & 0 \\ 0 & 0 \end{pmatrix} + \begin{pmatrix} \bar{X}' \\ -I \end{pmatrix} \cdot L_{xx}^{-1} \cdot (\bar{X} - I)$$

$$= \begin{pmatrix} \frac{1}{n} + \bar{X}' L_{xx}^{-1} \bar{X} - \bar{X}' L_{xx}^{-1} \\ -L_{xx}^{-1}\bar{X} \qquad L_{xx}^{-1} \end{pmatrix}, \tag{9}$$

式中：

$$L_{xx}^{-1} = X'(I - \frac{1}{n}J)X, \quad \bar{X} = \frac{1}{n}X'1,$$

$$I = \begin{pmatrix} 1 & & \\ & 1 & \\ & & \ddots \\ & & & 1 \end{pmatrix}, \quad J = \begin{pmatrix} 1 & 1 & \cdots & 1 \\ \cdots & & & \\ 1 & 1 & \cdots & 1 \end{pmatrix}, \quad 1 = \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix} \tag{10}$$

记 $\beta = \begin{pmatrix} \beta_0 \\ B \end{pmatrix}$，相应的最小二乘解为：

$$\begin{pmatrix} \hat{\beta} \\ \hat{B} \end{pmatrix} = \left(\begin{pmatrix} 1' \\ X' \end{pmatrix} \cdot (1X)\right)^{-1} \cdot \begin{pmatrix} 1' \\ X' \end{pmatrix} \cdot Y$$

$$= \begin{pmatrix} \frac{1}{n} + \bar{X}' L_{xx}^{-1} \bar{X} - \bar{X}' L_{xx}^{-1} \\ -L_{xx}^{-1}\bar{X}' \qquad L_{xx}^{-1} \end{pmatrix} \cdot \begin{pmatrix} 1'Y \\ X'Y \end{pmatrix} \tag{11}$$

令：$\bar{Y} = \frac{1}{n}Y'1, \quad L_{yy} = Y'(I - \frac{1}{n}J)Y,$ (12)

$$L_{xy} = L_{yx} = X'(I - \frac{1}{n}J)Y$$

其最小二乘估计为：

$$\begin{pmatrix} \hat{\beta} \\ \hat{B} \end{pmatrix} = \begin{pmatrix} (\frac{1}{n} + \bar{X}' L_{xx}^{-1} \bar{X}) \cdot n\bar{Y}' - \bar{X}' L_{xx}^{-1} \bar{X} Y \\ -L_{xx}^{-1} \bar{X}(n\bar{Y}') + L_{xx}^{-1} X'Y \end{pmatrix}$$

$$= \begin{pmatrix} \bar{Y}' - \bar{X} L_{xx}^{-1}(X'Y - n\bar{X}'\bar{Y}') \\ L_{xx}^{-1}(X'Y - n\bar{X}'\bar{Y}') \end{pmatrix}$$

$$= \begin{pmatrix} \bar{Y}' - \bar{X}' L_{xx}^{-1} L_{xy} \\ L_{xx}^{-1} L_{xy} \end{pmatrix}$$

$$= \begin{pmatrix} \bar{Y} - \bar{X}' \hat{B} \\ L_{xx}^{-1} L_{xy} \end{pmatrix} \tag{13}$$

即：$\hat{\beta}_0 = \bar{Y} - \bar{X}' \hat{B}$，$\hat{B} = L_{xx}^{-1} L_{xy}$ (14)

回归方程为：$\hat{Y} = (1X) \cdot \begin{pmatrix} \hat{\beta}_0 \\ \hat{B} \end{pmatrix} = 1\hat{\beta}_0 + X \cdot \hat{B}$

$$= 1\bar{Y}' + (I + \frac{1}{n}J) X \hat{B} \tag{15}$$

残差为：$\hat{\varepsilon}' \cdot \hat{\varepsilon} = (Y - \hat{Y})' \cdot (Y - \hat{Y})$

$$= L_{yy} - L_{yx} L_{xx}^{-1} L_{xy} \tag{16}$$

显然，多变量对多变量线性回归的最小二乘解与常规多元线性回归的最小二乘解形式完全一致。

3 计算实施方法

在常规多元线性回归中，求解回归系数 $b_i (i = 1, 2, \cdots, p)$ 的正规方程组为：

$$\sum_{j=1}^{p} l_{ij} b_j = l_{iy}, \quad (i = 1, 2, \cdots, p) \tag{17}$$

式中：$l_{ij} = \sum_{k=1}^{n} (x_{ik} - \bar{x}_i)(x_{jk} - \bar{x}_j)$，

$$l_{iy}=\sum_{k=1}^{n}(x_{ik}-\bar{x}_i)(y_k-\bar{y})$$

$\bar{x}_i$和$\bar{y}$分别为各自变量和因变量的样本均值，在对正规方程组求解求得$b_i(i=1,2,\cdots,p)$后，再回求回归方程常数项：

$$b_0=\bar{y}-\sum_{i=1}^{p}b_i\bar{x}_i。\tag{18}$$

在实际求解时，为了提高计算精度和求解有关统计检验参数的需要，利用求解由标准化回归系数和相关系数组成的标准化正规方程组求得有关结果。此时：

$$\sum_{j=1}^{p}r_{ij}\cdot b_j^{*}=r_{iy}\qquad(i=1,2,\cdots,p)\tag{19}$$

其中b_j^{*}为标准化回归系数，与常规回归系数b_j的关系为：

$$b_j=b_j^{*}\sqrt{l_{yy}}\Big/\sqrt{l_{jj}}\tag{20}$$

通过对包含了y因变量相关系数的相关系数矩阵：

$$R=\begin{pmatrix}r_{11}&r_{12}&\cdots&r_{1p}&r_{1y}\\&r_{22}&\cdots&r_{2p}&r_{2y}\\&&&r_{pp}&r_{py}\\&&&&r_{yy}\end{pmatrix}\tag{21}$$

进行求解求逆紧凑方法迭代运算，最终求得包含了自变量相关系数矩阵的逆矩阵和标准化回归系数的新矩阵：

$$R'=\begin{pmatrix}r'_{11}&r'_{12}&\cdots&r'_{1p}&b_1^{*}\\&r'_{22}&\cdots&r'_{2p}&b_2^{*}\\&&&r'_{pp}&b_p^{*}\\&&&&r'_{yy}\end{pmatrix}\tag{22}$$

对角线最右下角元素r'_{yy}即为在标准化条件下（l_{yy}取为1）的残差平方和。

对于本文介绍的多变量对多变量线性回归，其起始矩阵为一大的实对称矩阵，它的每一分块矩阵：左上角分块和右下角分块分别为自变量本身与因变量本身的相关系数对称矩阵；左下角和右上角同为对称的自变量与因变量之间的相关系数矩阵：

$$R=\left(\begin{array}{cccc|cccc}r_{11}&r_{12}&\cdots&r_{1p}&r_{1y_1}&r_{1y_2}&\cdots&r_{1y_m}\\&r_{22}&\cdots&r_{2p}&r_{2y_1}&r_{2y_2}&\cdots&r_{2y_m}\\&&\ddots&&&\cdots&&\\&&&r_{pp}&r_{py_1}&r_{py_2}&\cdots&r_{py_m}\\\hline&&&&r_{y_1y_1}&r_{y_1y_2}&\cdots&r_{y_1y_m}\\&&&&&r_{y_2y_2}&\cdots&r_{y_2y_m}\\&&&&&&\ddots&\\&&&&&&&r_{y_my_m}\end{array}\right)\tag{23}$$

经过 p 次求解求逆方案的迭代运算，求得自变量相关系数逆矩阵，标准化回归系数和剩余残差平方和为：

$$R' = \begin{array}{c|c} \begin{matrix} r'_{11} & r'_{12} & \cdots & r'_{1p} \\ & r'_{22} & \cdots & r'_{2p} \\ & & \ddots & \\ & & & r'_{pp} \end{matrix} & \begin{matrix} b^{*}_{1y_1} & b^{*}_{1y_2} & \cdots & b^{*}_{1y_m} \\ b^{*}_{2y_1} & b^{*}_{2y_2} & \cdots & b^{*}_{2y_m} \\ \cdots & & & \\ b^{*}_{py_1} & b^{*}_{py_2} & \cdots & b^{*}_{py_m} \end{matrix} \\ \hline \begin{matrix} r'_{y_1y_1} & & & \\ & r'_{y_2y_2} & & \\ & & \ddots & \\ & & & r'_{y_my_m} \end{matrix} & \end{array} \tag{24}$$

$b^{*}_{iy_j}$ 表示第 i 个自变量在第 j 个因变量的回归方程中的标准化回归系数。由式(20)反求 b_{iy_j}，式（18）回求常数项，从而组成回归方程。

计算每一个因变量与自变量之间的复相关系数，每一个自变量在回归方程中所起作用大小的偏回归平方和及其 t 检验值，以及作为预测控制精度衡量指标的剩余标准差，其方法均与常规的多元回归分析方法相同。

4 区域研究运用实例

解放以来，我国人民的生活水平有了极大的提高。地处我国经济腹地中部的湖北省，人民生活水平同样有了极大提高。衡量人民生活水平状态有众多指标，但包括吃、穿、烧、用在内的社会消费品零售额可从一个重要的侧面反映着人民生活水平的现状。研究人民生活水平的历史、现状和演变是区域研究的一个重要内容，它能较为全面地反映区域内人与自然，人与社会，人与经济之间的相互制约关系。我们在研究湖北省人民生活水平（全省吃穿用消费品零售额）和生产与经济相互关系时，利用多变量对多变量线性回归为工具，选择了如下变量：

自变量：1 全省人平国民收入：x_1　2 工业农业总产值生产指数：x_2　3 人口自然增长率：x_3　4 全省社会商品零售总额：x_4　5 零售物价总指数：x_5

自变量x_2，x_5均以1950年为基准（100）。

因变量：1 全省“吃”消费品零售额：y_1

2 全省“穿”消费品零售额：y_2

3 全省“用”消费品零售额：y_3

为了计算上的方便，已将“烧”消费品零售额归并入y_1。

所有变量求得的相关系数矩阵为（样本年限为1956～1976）：

$$R=\begin{pmatrix}
1 & 0.9900 & -0.4367 & 0.9934 & 0.6834 & 0.9927 & 0.9889 & 0.9809\\
 & 1 & -0.4799 & 0.9958 & 0.6608 & 0.9886 & 0.9909 & 0.9568\\
 & & 1 & -0.4555 & -0.0156 & -0.4086 & -0.4667 & -0.3927\\
 & & & 1 & 0.6919 & 0.9959 & 0.9953 & 0.9751\\
 & & & & 1 & 0.7116 & 0.6527 & 0.7341\\
 & & & & & 1 & 0.9870 & 0.9833\\
 & & & & & & 1 & 0.9679\\
 & & & & & & & 1
\end{pmatrix}$$

经过求解有逆紧凑方法迭代运算，求得的含有自变量相关系数逆矩阵，标准化回归系数和标准化剩余平方和的矩阵：

$$R'=\begin{pmatrix}
80.04849 & -10.21177 & -2.49601 & -71.51646 & 1.48610 & 0.22292 & -0.00435 & 1.08438\\
 & 150.72034 & 2.75364 & -143.18082 & 6.49248 & -0.23395 & -0.30497 & -1.73833\\
 & & 1.72946 & 1.12288 & -0.86376 & 0.04141 & 0.00695 & -0.03199\\
 & & & 222.23181 & -10.25644 & 1.02042 & 1.36840 & 1.58213\\
 & & & & 2.77713 & 0.00846 & -0.08948 & -0.04655\\
 & & & & & 0.00465 & & \\
 & & & & & & 0.00619 & \\
 & & & & & & & 0.01010
\end{pmatrix}$$

由标准化回归系数回求所得的一般回归系数与常数项为表1。

表1　回归系数与常数项

	x_1	x_2	x_3	x_4	x_5	b_0
y_1	0.02869	−0.00816	0.11989	0.47048	0.01528	−6.58862
y_2	−0.00022	−0.00421	0.00796	0.24964	−0.06396	7.88402
y_3	0.09633	−0.04185	−0.06394	0.50359	0.05806	−17.02462

因变量y_1,y_2,y_3与自变量$x_1,\cdots,x_5$回归方程的复相关系数，剩余标准差分别为表2。

表2　回归方程的复相关系数剩余标准差

	r	s
y_1	0.9977	1.6686
y_2	0.9969	0.7617
y_3	0.9949	1.6976

拟合优化度很高，利用回归方程求得的拟合值误差全部在两倍剩余标准差的误差范围内。

至于各个自变量在回归方程中的作用显著性T检验，利用由相关系数逆矩阵元素推得的离差矩阵的逆矩阵元素，计算得t值为表3。

表 3　回归方程显著性T检验结果

	t_1	t_2	t_3	t_4	t_5
y_1	2.12	1.41	2.31	5.02	0.37
y_2	0.83	2.06	0.34	5.83	3.41
y_3	6.01	7.01	1.21	5.28	1.39

置信度 $\alpha=0.05$时，$n-p-1=25$的 $t_{0.05}(25)=2.060$。显然，每个变量在方程中的显著性是不相同的。

利用建立的回归方程，以1977年样本资料作预测检验，由5个自变量求得的预测值，实际资料，误差如表4。

表 4　预测值与实际值的比较

	预测 $\hat{y}$	实际 y	$y-\hat{y}$	$(y-\hat{y})/y\times100$
y_1	113.6000	114.250	0.65	0.57
y_2	43.4093	40.100	−3.3093	−8.25
y_3	81.7701	79.710	−2.0601	−2.58

预测有相当高的精度。这充分显示了多变量对多变量线性回归方法在区域研究中的潜在用途。

参 考 文 献

1　Anderson, T.W.,An Introduction to Multivariate Statistical Analysis. John Wiley & Sons, 1958

2　中国科学院计算中心概率统计组。概率统计计算。北京：科学出版社，1979

3　张尧庭，方开泰。多元统计分析引论。北京：科学出版社，1982

4　罗积玉，邢　瑛。经济统计分析方法及预测。北京：清华大学出版社，1987

A LINEAR REGRESSIVE METHOD OF MULTIVARIATE TO MULTIVARIATE AND ITS APPLICATION IN REGIONAL RESEARCH

Liu Miaolong

(Institute of Regional Planning)

Abstract

The linear regressive method of multivariate to multivariate, which is widely used in regional research, is discussed. A model, solution of least square and process of calculation are given. This paper, taking the living standard of the people of Hubei province and its social and economic conditions for example, introduces the practical application of the method.

Key Words　Multivariate, Linear Regressive, Regional Research.