合理进行多元分析——多元多重线性回归分析

胡纯严[1],胡良平[1,2*]

(1. 军事科学院研究生院,北京 100850;

2. 世界中医药学会联合会临床科研统计学专业委员会,北京 100029

*通信作者:胡良平,E-mail:lphu927@163.com)

【摘要】 本文目的是介绍与多元多重线性回归分析有关的基本概念、计算方法、两个实例以及SAS实现。基本概念包括多元多重回归分析、普通最小二乘法、偏最小二乘法、主成分分析、典型相关分析;计算方法涉及准备数据和实施步骤;两个实例的资料分别是“汉族男性学生的身体形态学指标与功能指标的测定结果”以及“两组受试者身体形态与健康状况指标的测定结果”;借助SAS,对两个实例的数据分别进行了多元多重线性回归分析,并对SAS输出结果做出了解释。

【关键词】 多元多重线性回归分析;普通最小二乘法;偏最小二乘法;主成分分析;典型相关分析

中图分类号:R195.1 文献标识码:A doi:10.11886/scjsws20230830001

Reasonably carry out multivariate analysis: multivariate multiple linear regression analysis

Hu Chunyan[1], *Hu Liangping*[1,2*]

(1. *Graduate School, Academy of Military Sciences PLA China, Beijing* 100850, *China*;

2. *Specialty Committee of Clinical Scientific Research Statistics of World Federation of Chinese Medicine Societies, Beijing* 100029, *China*

**Corresponding author: Hu Liangping, E-mail: lphu927@163.com*)

【Abstract】 The purpose of this article was to introduce the basic concepts, calculation methods, two examples and SAS implementation related to the multivariate multiple linear regression analysis. Basic concepts included multivariate multiple regression analysis, ordinary least squares, partial least squares, principal component analysis and canonical correlation analysis. The calculation methods involved data preparation and implementation steps. The data in the two examples were "measurement results of body morphology indicators and functional indicators of Han male students" and "measurement results of physical shape and health status indicators of two groups of subjects". With the help of SAS software, the multivariate multiple linear regression analysis was carried out on the data in the two examples, and an explanation was made for the output results of SAS.

【Keywords】 Multivariate multiple linear regression analysis; Ordinary least squares method; Partial least squares method; Principal component analysis; Canonical correlation analysis

在许多实际问题中,往往同时存在多个因变量和多个自变量。若基于每个因变量构建一个一元多重线性回归模型,不仅效率较低,还割裂了因变量之间可能存在的某些内在联系。为了更全面且真实地揭示一组因变量随另一组自变量变化而变化的依赖关系,需采用偏最小二乘法,并构建全部因变量关于一组自变量的多重线性回归方程组。本文在介绍有关基本概念的基础上,介绍基于偏最小二乘法实现多元多重线性回归分析的具体方法,并通过两个实例演示如何使用SAS实现计算的全过程。

1 基本概念

1.1 多元多重回归分析

在通常的回归分析中,涉及两类可观测的变量(也称为显变量),即因变量与自变量。当因变量的个数P=1时,自变量的个数m可以有以下两种情况:其一,当m=1时,称为一元一重回归分析,最常见的是直线回归分析;其二,当m>1时,称为一元多重回归分析,最常见的是一元多重线性回归分析,简称为多重线性回归分析。当因变量的个数P>1时,自变量的个数m可以有以下两种情况:其一,当m=1时,称为P元一重回归分析,最常见的是P元一重线性回归分析;其二,当m>1时,称为P元多重回归分析,最常见的是P元多重线性回归分析,简称为多元多重线性回归分析。

1.2 普通最小二乘法

在一定的条件下,人们希望研究一个因变量随一个或多个自变量变化而变化的依赖关系时,常事先做出一些“假定”,并且当自变量取不同值时,依据此“假定”推算出因变量的取值(称为因变量的理

论值)。构造各点(即自变量取不同数值)上因变量的观测值与其理论值之差的平方和,关键是使此"平方和"成为某个或某些待定的"未知参数"的函数。于是,将那些"未知参数"视为"自变量",再设法使此"平方和"取得最小值。这样求得的"解"就是关于"未知参数"的计算公式。用这样的思路导出未知参数计算公式的方法被称为普通最小二乘法或普通最小平方法[1]。这里的"普通"是指每个观测点在计算时被认为是同等重要的,即每个观测点的"权重"都是"1"。

1.3 偏最小二乘法

在估计多元多重线性回归模型中的参数时,采用多步来实现。其中,每一步只要涉及模型参数估计(模型从含一个"自变量"开始,每次递增一个,且"自变量"是从原始自变量组与因变量组提取的主成分变量,主成分变量按其贡献由大到小逐次引入),也都采用普通最小二乘法。基于多次运用普通最小二乘法导出的最终回归模型中的系数,被称为偏最小二乘估计值[2];经过多步运用普通最小二乘法实现多元多重线性回归分析的方法被称为偏最小二乘法。

1.4 主成分分析

在实施偏最小二乘分析的过程中,涉及主成分分析。主成分分析就是基于单组设计m元定量资料构造协方差矩阵或相关矩阵,求前述提及的两种矩阵之一的特征值及其特征向量。于是,将观测变量组成的向量与第i个(i=1,2,…,m)特征向量求内积(本质上是一个表达式),赋值给一个隐变量P_i,则它就是第i个(i=1,2,…,m)主成分变量。在统计学上,基于求得的主成分变量,并结合其表达式中各系数的绝对值和正负号,给出合理的专业解释的过程,称为主成分分析[3]。

1.5 典型相关分析

在实施偏最小二乘分析的过程中,还涉及典型相关分析。典型相关分析就是对由两组定量变量产生的多对主成分变量(注:在典型相关分析中,称这些成对的主成分变量为典型变量对)进行相关分析[4]。

2 计算方法

2.1 准备数据

偏最小二乘法是估计多个因变量对多个自变量建立多元多重线性回归方程中回归参数的方法。假设在某实际问题中,有m个因变量($y_1,y_2,\cdots,y_m$)和k个自变量($x_1,x_2,\cdots,x_k$)。对具有同质性的n个个体同时观测这两组变量的取值,得到两个数据矩阵或数据表,见式(1)。

$$y=\begin{pmatrix} y_{11} & y_{12} & \cdots & y_{1m} \\ y_{21} & y_{22} & \cdots & y_{2m} \\ \vdots & \vdots & \cdots & \vdots \\ y_{n1} & y_{n2} & \cdots & y_{nm} \end{pmatrix}_{n\times m} \qquad x=\begin{pmatrix} x_{11} & x_{12} & \cdots & x_{1k} \\ x_{21} & x_{22} & \cdots & x_{2k} \\ \vdots & \vdots & \cdots & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{nk} \end{pmatrix}_{n\times k} \tag{1}$$

依据这些观测数据,最终目的是建立如下回归方程组,见式(2)。

$$\begin{cases} y_1 = b_{10} + b_{11}x_1 + b_{12}x_2 + \cdots + b_{1k}x_k \\ y_2 = b_{20} + b_{21}x_1 + b_{22}x_2 + \cdots + b_{2k}x_k \\ \qquad\vdots \\ y_m = b_{m0} + b_{m1}x_1 + b_{m2}x_2 + \cdots + b_{mk}x_k \end{cases} \tag{2}$$

式(2)中的回归系数并非是采用一组公式一次性求出的,而是经过下列多个步骤求出的[5]。

2.2 实施步骤

第一步:提取第一对主成分变量。基于主成分分析方法并结合典型相关分析方法,分别求出自变量$x_1,x_2,\cdots,x_k$组中的第一主成分变量U_1与因变量$y_1,y_2,\cdots,y_m$组中的第一主成分变量V_1,分别见式(3)、式(4)。

$$U_1 = u_{11}x_1 + u_{12}x_2 + \cdots + u_{1k}x_k \tag{3}$$

$$V_1 = v_{11}y_1 + v_{12}y_2 + \cdots + v_{1m}y_m \tag{4}$$

求得的第一对主成分变量(U_1,V_1)应满足以下两个条件:其一,U_1和V_1应尽可能多地携带它们各自数据表中的变异信息;其二,U_1和V_1之间的相关程度能够达到最大。

第二步:建立($k+m$)个一元一重线性回归方程(基于普通最小二乘法),自变量均为U_1,见式(5)。

$$\begin{cases} x_1 = a_{11}U_1 + e_1 \\ x_2 = a_{12}U_1 + e_2 \\ \qquad\vdots \\ x_k = a_{1k}U_1 + e_k \\ y_1 = b_{11}U_1 + \varepsilon_1 \\ y_2 = b_{12}U_1 + \varepsilon_2 \\ \qquad\vdots \\ y_m = b_{1m}U_1 + \varepsilon_m \end{cases} \tag{5}$$

式(5)中各方程中的最后一项均为残差。若式(5)中最后m个方程式中的残差的绝对值都几乎为0,则认为仅采用第一个主成分变量就已满足精度

要求了，于是，停止抽取新的主成分。

第三步：建立残差矩阵取代原先的数据矩阵。

若式(5)中最后 m 个方程式中的残差的绝对值并非都接近于0，需要用以下残差矩阵分别取代式(1)中的两个原始数据矩阵，见式(6)，再重复前文第一步的做法。

$$\Xi = \begin{pmatrix} \varepsilon_{11} & \varepsilon_{12} & \cdots & \varepsilon_{1m} \\ \varepsilon_{21} & \varepsilon_{22} & \cdots & \varepsilon_{2m} \\ \vdots & \vdots & \cdots & \vdots \\ \varepsilon_{n1} & \varepsilon_{n2} & \cdots & \varepsilon_{nm} \end{pmatrix}_{n \times m} \quad E = \begin{pmatrix} e_{11} & e_{12} & \cdots & e_{1k} \\ e_{21} & e_{22} & \cdots & e_{2k} \\ \vdots & \vdots & \cdots & \vdots \\ e_{n1} & e_{n2} & \cdots & e_{nk} \end{pmatrix}_{n \times k} \tag{6}$$

第四步：假定经过 r 步之后［注：设 r 为式(1)中数据矩阵 x 的秩，即 $r \le \min(n-1,k)$］，对应残差矩阵中所有元素的绝对值都几乎为0，此时，就彻底结束了上述的分解过程。于是，可写出与式(5)类似且每个方程式中都包含 r 个主成分变量的表达式，见式(7)。

$$\begin{cases} x_1 = a_{11}U_1 + a_{21}U_2 + \cdots + a_{r1}U_r + e_1 \\ x_2 = a_{12}U_1 + a_{22}U_2 + \cdots + a_{r2}U_r + e_2 \\ \vdots \\ x_k = a_{1k}U_1 + a_{2k}U_2 + \cdots + a_{rk}U_r + e_k \\ y_1 = b_{11}U_1 + b_{21}U_2 + \cdots + b_{r1}U_r + \varepsilon_1 \\ y_2 = b_{12}U_1 + b_{22}U_2 + \cdots + b_{r2}U_r + \varepsilon_2 \\ \vdots \\ y_m = b_{1m}U_1 + b_{2m}U_2 + \cdots + b_{rm}U_r + \varepsilon_m \end{cases} \tag{7}$$

式(7)中各方程中的最后一项均为残差；各 $U_i(i=1,2,\cdots,r)$ 都有形如式(3)的表达式，即都是自变量组中全部自变量的线性组合。

第五步：用原始的自变量取代主成分变量 U。在式(7)中的最后 m 个方程式中，针对各 U_i，用类似式(3)等号后面的表达式取而代之，再将各相同自变量 $x_j(j=1,2,\cdots,k)$ 前的系数合并，就可获得式(2)那样的最终结果，它就被称为基于偏最小二乘法拟合的多元多重线性回归方程组。

3 实例与SAS实现

3.1 问题与数据结构

3.1.1 2个实际问题及数据

【例1】研究19~22岁汉族男性学生的身体形态学指标与功能指标之间的关系，从特定的总体中随机抽取并测量了28名受试者，具体数据见表1[6]。试以功能指标为因变量，以形态学指标为自变量，采用偏最小二乘法建立它们之间的多元多重线性回归方程。

【例2】测量15名受试者的身体形态和健康状况指标，资料见表2[7]。第一类资料是身体形态变量，包括年龄、体重、日吸烟量和胸围；第二类是健康状况变量，包括脉搏、收缩压和舒张压。构建健康状况与身体形态这两类变量之间的依赖关系的多元多重线性回归方程。

3.1.2 对数据结构的分析

例1和例2的资料都属于单组设计多元定量资料。结合实际问题可知，每个实例中的全部定量资料按性质可以被分为两类，例如，一类为反映人体形态学方面的指标，另一类为反映功能方面的指标。

表1 汉族男性学生的身体形态学指标与功能指标的测定结果

Table 1 Measurement results of body morphological and functional indicators of Han male students

编 号	各形态学指标的取值						各功能指标的取值				
	身高（cm）	坐高（cm）	体重（kg）	胸围（cm）	肩宽（cm）	骨盆径（cm）	脉搏（times/min）	收缩压（mmHg）	原舒张压（mmHg）	舒张末期压（mmHg）	肺活量（mL）
1	173.28	93.62	60.10	86.72	38.97	27.51	75.3	117.4	74.6	61.8	4 508
2	172.09	92.83	60.38	87.39	38.62	27.82	76.7	120.1	77.1	66.2	4 469
…	…	…	…	…	…	…	…	…	…	…	…
28	168.99	91.52	55.11	86.23	38.30	27.14	77.7	113.3	72.1	52.8	4 238

表2 两组受试者身体形态与健康状况指标的测定结果

Table 2 Measurement results of physical shape and health status indicators of two groups of subjects

编 号	年龄(X_1)	体重(X_2)	日吸烟量(X_3)	胸围(X_4)	脉搏(Y_1)	收缩压(Y_2)	舒张压(Y_3)
1	25	125	30	83.5	70	130	85
2	26	131	25	82.9	72	135	80
…	…	…	…	…	…	…	…
14	48	139	50	81.6	85	150	95
15	45	140	55	88.0	88	160	95

3.1.3 创建SAS数据集

分析例1资料，设所需SAS数据步程序如下：

```
data a1;
input X1-X6 Y1-Y5;
cards;
173.28 93.62 60.10 86.72 38.97 27.51 75.3
117.4 74.6 61.8 4508
172.09 92.83 60.38 87.39 38.62 27.82 76.7
120.1 77.1 66.2 4469
（此处省略了多行数据）
168.15 91.50 54.56 84.81 38.44 27.38 77.5
117.4 75.3 63.6 4039
168.99 91.52 55.11 86.23 38.30 27.14 77.7
113.3 72.1 52.8 4238
;
run;
```

分析例2资料，设所需SAS数据步程序如下：

```
data a2;
input X1-X4 Y1-Y3;
cards;
25 125 30 83.5 70 130 85
26 131 25 82.9 72 135 80
28 128 35 88.1 75 140 90
29 126 40 88.4 78 140 92
27 126 45 80.6 73 138 85
32 118 20 88.4 70 130 80
31 120 18 87.8 68 135 75
34 124 25 84.6 70 135 75
36 128 25 88.0 75 140 80
38 124 23 85.6 72 145 86
41 135 40 86.3 76 148 88
46 143 45 84.8 80 145 90
47 141 48 87.9 82 148 92
48 139 50 81.6 85 150 95
45 140 55 88.0 88 160 95
;
run;
```

3.2 用SAS实现统计分析

3.2.1 分析例1的资料

设所需SAS过程步程序如下[8]：

```
proc pls data=a1 censcale cv=one cvtest(pval=0.05 stat=press)
method=pls details;
model Y1-Y5=X1-X6 / solution;
run;
```

【SAS输出结果及解释】每个因变量的平均值和标准差的计算结果见表3。

表3 各因变量的平均值和标准差

Table 3 Mean and standard deviation of each dependent variable

因变量	平均值	标准差
Y_1	75.596	1.641
Y_2	116.011	2.559
Y_3	73.686	2.283
Y_4	60.114	5.579
Y_5	4 251.357	161.173

由以上结果可知，在对5个因变量进行标准化时，对每个个体而言，其每个变量都需要减去平均值（第二列）、除以标准差（第三列）。每个自变量的平均值和标准差的计算结果见表4。

表4 各自变量的平均值和标准差

Table 4 Mean and standard deviation of each independent variable

自变量	平均值	标准差
X_1	170.370	1.437
X_2	92.199	0.690
X_3	57.694	1.723
X_4	86.067	1.336
X_5	38.478	0.454
X_6	27.171	0.384

由表4可知，在对6个自变量进行标准化时，对每个个体而言，其每个变量都需要减去平均值（第二列）、除以标准差（第三列）。基于预测残差平方和（prediction error of square sum，PRESS）的数值进行比较，以判断提取多少个公因子的检验结果见表5。

表5 基于PRESS对提取的因子数交叉验证的结果

Table 5 Cross validation results of extracted factor numbers based on PRESS

提取的因子数	PRESS	概率>PRESS
0	1.037	0.002
1	0.911	1.000
2	0.945	0.430
3	0.989	0.397
4	1.025	0.257
5	1.003	0.183
6	1.064	0.114

由表5可知，基于“PRESS”进行检验，以提取一个公因子（表5第二行）为基准，当提取公因子数≥2时，差异均无统计学意义。由此可知，对本例资料而言，提取一个公因子即可。当PRESS取最小值时，对应的公因子数见表6。由表6可知，当抽取的主成分个数的最小值为1时，对应的预测残差均方

根的最小值为0.911，在决定抽取最小主成分个数时所采用的显著性水平为P>0.05［即当对抽取的两个不同个数主成分构造出的两个模型（其中一个为使PRESS取最小值的模型，令其为对照模型）进行差异性检验时，若P>0.05时，则认为它们之间的差异无统计学意义］。

表6　PRESS取最小值时对应的公因子数

Table 6　Number of common factors corresponding to the minimum value of PRESS

统计量	数　值
最小PRESS	0.911
最小化因子数	1
最小因子数（P>0.05）	0

在变量标准化的前提下，各多重线性回归方程中与各自变量对应的回归系数的估计结果见表7。

表7　基于变量标准化变换后的多重线性回归方程中回归系数的估计结果

Table 7　Estimation results of regression coefficients in multiple linear regression equations based on the variable normalization transformation

自变量	Y_1	Y_2	Y_3	Y_4	Y_5
Intercept	0.000	0.000	0.000	0.000	0.000
X_1	−0.086	0.122	0.166	0.137	0.212
X_2	−0.086	0.121	0.164	0.135	0.210
X_3	−0.097	0.137	0.186	0.153	0.239
X_4	−0.065	0.091	0.124	0.102	0.158
X_5	−0.031	0.043	0.059	0.048	0.075
X_6	−0.041	0.057	0.078	0.064	0.100

表7是用标准化的全部自变量线性表达每一个标准化的因变量的表达式中的回归系数。现以表7第二列为例，写出表达式，见式(8)。

$$y_1^* = -0.086x_1^* - 0.086x_2^* - \cdots - 0.041x_6^* \qquad (8)$$

其他4个表达式可按同样方法写出，此处从略。注意：式(8)中，变量右上角带"*"号，表明这些变量属于经标准化变换的变量。

在原始变量的前提下，各个多重线性回归方程中与各自变量对应的回归系数的估计结果见表8。

表8　基于原始变量的多重线性回归方程中回归系数的估计结果

Table 8　Estimation results of regression coefficients in multiple linear regression equations based on the original variables

自变量	Y_1	Y_2	Y_3	Y_4	Y_5
Intercept	132.289	−8.918	−77.613	−244.224	−9 420.005
X_1	−0.099	0.218	0.264	0.530	23.823
X_2	−0.203	0.448	0.543	1.092	49.049
X_3	−0.092	0.204	0.247	0.497	22.304
X_4	−0.079	0.175	0.211	0.425	19.107
X_5	−0.110	0.243	0.295	0.593	26.618
X_6	−0.173	0.382	0.462	0.930	41.763

表8是用原始的全部自变量线性表达每一个原始的因变量的表达式中的回归系数。现以表8第二列为例，写出表达式，见式(9)。

$$\hat{y}_1 = 132.289 - 0.099x_1 - 0.203x_2 - \cdots - 0.173x_6 \quad (9)$$

其他4个表达式可按同样方法写出，此处从略。注意：式(9)中，各变量代表未经过标准化变换的原始变量。

把5个多重线性回归方程全部写出后，就实现了研究者分析此资料的最终目的：即基于偏最小二乘法构建五元六重线性回归方程组（注：这是基于抽取一对主成分变量产生的结果）。

3.2.2　分析例2的资料

设所需要的SAS过程步程序如下[8]：

```
proc pls data=a2 censcale cv=one cvtest (pval=0.05 stat=press)
method=pls details;
model Y1-Y3=X1-X4 / solution;
run;
```

【SAS输出结果及解释】因篇幅所限，仅输出最终的计算结果。在原始变量的前提下，各个多重线性回归方程中与各自变量对应的回归系数的估计结果见表9。

表9　基于原始变量的多重线性回归方程中回归系数的估计结果

Table 9　Estimation results of regression coefficients in multiple linear regression equations based on the original variables

自变量	Y_1	Y_2	Y_3
Intercept	21.369	71.488	33.800
X_1	0.227	0.292	0.218
X_2	0.261	0.336	0.251
X_3	0.183	0.236	0.176
X_4	0.068	0.087	0.065

表9是用原始的全部自变量线性表达每一个原始因变量的表达式中的回归系数。现以表9第二列为例，写出表达式，见式(9)。

$$\hat{y}_1 = 21.369 + 0.227x_1 + 0.261x_2 + 0.183x_3 + 0.068x_4 \quad (9)$$

其他两个表达式可按同样方法写出，此处从略。把3个多重线性回归方程全部写出后，就实现了研究者分析此资料的最终目的：即基于偏最小二乘法构建三元四重线性回归方程组（注：这是基于抽取一对主成分变量产生的结果）。

4　讨论与小结

4.1　讨论

采用proc pls过程可以对k(k>1)个因变量同时

建立$k(k>1)$个多重线性回归方程；若采用proc reg过程，则需要分别对每个因变量建立一元多重线性回归方程，共需进行$k(k>1)$次。然而，采用前述两种策略构建的多重线性回归方程组不是完全相同的，它们之间的区别在于：采用proc pls过程可以消除变量之间存在的多重相关性以及自变量之间的多重共线性对回归分析结果的影响；而通过proc reg过程多次构建的多个多重线性回归方程之间是彼此互相独立的，也无法消除自变量之间可能存在的多重共线性对回归分析结果的影响(通常需借助主成分回归分析或岭回归分析等方法来间接实现)。

虽然基于proc pls过程可以得到多元多重线性回归方程组，但经过多步计算得到的多元多重线性回归方程组中的每个回归方程以及各回归方程中各回归系数是否都有统计学意义，并非十分清楚，需要通过其他途径予以解决，因篇幅所限，此处从略；而采用proc reg过程时，在过程步程序中增加一个语句“mtest X1-X6；”，就可采用多元方差分析方法同时检验多元多重线性回归方程组是否有统计学意义[8]。

在构建多重线性回归模型方面，基于偏最小二乘法比基于普通最小二乘法具有更多的优点，例如，在对自变量进行信息综合时，不但考虑了要更好地概括自变量系统中的信息，而且注重要求所提取的成分必须对因变量也有较好的解释；无需剔除任何子变量和样本点，具有简单稳健、易于定性解释、预测精度较高等优点，尤其当自变量个数较多、样本量少时，其数据探索性分析更有效；此法对系统信息和噪声有良好的辨识能力，能使模型结果对实测变量的物理成因解释更合理[9]。

4.2 小结

本文介绍了与多元多重回归分析有关的基本概念、计算方法、两个实例以及使用SAS实现计算的方法。基本概念包括多元多重回归分析、普通最小二乘法、偏最小二乘法、主成分分析、典型相关分析；计算方法涉及准备数据和实施步骤；两个实例的资料分别是“汉族男性学生的身体形态学指标与功能指标的测定结果”和“两组受试者身体形态与健康状况指标的测定结果”；借助SAS，对两个实例的数据分别进行了多元多重线性回归分析，并对SAS输出结果做出了解释。

参考文献

[1] Neil H. Timm, Tammy A. Mieczkowski. General linear models: univariate & multivariate theory and applications using SAS software[M]. Cary, NC: SAS Institute Inc, 1997: 42-43.

[2] Armitage P, Colton T. Encyclopedia of biostatistics [M]. 2nd edition. New York: John Wiley & Sons, Inc, 2005: 4911-4913.

[3] Johnson RA, Wichern DW. 实用多元统计分析[M]. 6版. 北京: 清华大学出版社, 2008: 430-480.
Johnson RA, Wichern DW. Applied multivariate statistical analysis [M]. 6th edition. Beijing: Tsinghua University Press, 2008: 430-480.

[4] 王静龙. 多元统计分析[M]. 北京: 科学出版社, 2008: 333-345.
Wang JL. Multivariate statistical analysis [M]. Beijing: Science Press, 2008: 333-345.

[5] 高惠璇. 应用多元统计分析[M]. 北京: 北京大学出版社, 2005: 369-379.
Gao HX. Applied multivariate statistical analysis [M]. Beijing: Peking University Press, 2005: 369-379.

[6] 胡良平, 胡纯严, 鲍晓蕾. 应用数理统计[M]. 北京: 电子工业出版社, 2015: 174-184.
Hu LP, Hu CY, Bao XL. Applied mathematical statistics [M]. Beijing: Publishing House of Electronics Industry, 2015: 174-184.

[7] 胡良平. 面向问题的统计学: (3)试验设计与多元统计分析[M]. 北京: 人民卫生出版社, 2012: 84-104.
Hu LP. Problem-oriented statistics: (3) experimental design and multivariate statistical analysis [M]. Beijing: People's Medical Publishing House, 2012: 84-104.

[8] SAS Institute Inc. SAS/STAT®15.1 user's guide[M]. Cary, NC: SAS Institute Inc, 2018: 7581-7636, 8429-8612.

[9] 李卫东. 应用多元统计分析[M]. 北京: 北京大学出版社, 2008: 275-287.
Li WD. Applied multivariate statistical analysis [M]. Beijing: Peking University Press, 2008: 275-287.

(收稿日期:2023-08-30)
(本文编辑:陈 霞)