

# 合理进行多元分析——典型相关分析

胡纯严<sup>1</sup>, 胡良平<sup>1,2\*</sup>

(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】** 本文目的是介绍与典型相关分析有关的基本概念、计算方法、两个实例以及 SAS 实现。基本概念包括典型变量、典型相关系数、典型相关分析、典型变量个数的确定以及典型变量的性质; 计算方法涉及典型相关变量中系数的估计和典型相关系数的假设检验; 两个实例涉及的资料分别是“受试者的身体形态以及健康指标的测定结果”和“人力资源管理实践变量与组织绩效变量的测定结果”; 借助 SAS 对两个实例中的定量资料进行了典型相关分析, 并对输出结果给出了解释。

**【关键词】** 多元分析; 典型变量; 典型相关系数; 典型相关分析; 拉格朗日乘子法

中图分类号: R195.1

文献标识码: A

doi: 10.11886/scjsws20230605004

## Reasonably carry out multivariate analysis: canonical correlation analysis

Hu Chunyan<sup>1</sup>, Hu Liangping<sup>1,2\*</sup>

(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 the paper was to introduce the basic concepts, calculation methods, two examples and SAS implementation related to the canonical correlation analysis. Basic concepts included canonical variables, canonical correlation coefficients, canonical correlation analysis, determination of the number of canonical variables and properties of canonical variables. The calculation approaches involved the estimation of the coefficients in the canonical correlation variables and the hypothesis testing of the canonical correlation coefficients. The data involved in the two examples were measurement results of subjects' body shape and health indicators and measurement results of human resource management practice variables and organizational performance variables. With the help of SAS software, the canonical correlation analysis was carried out on the quantitative data in the two instances, and a reasonable explanation was given for the output results.

**【Keywords】** Multivariate analysis; Canonical variable; Canonical correlation coefficient; Canonical correlation analysis; Lagrange multiplier method

简单相关分析是通过计算两个变量之间的简单相关系数(根据资料的性质不同,通常有 Pearson 相关系数、Spearman 秩相关系数和 Kendall 秩相关系数)以及对相关系数是否为零的假设检验来完成的;复相关分析是研究一个变量与一组变量之间的相关性的方法。本文介绍一种研究两组变量之间相关性的多元统计分析方法,即典型相关分析。通过两个实例,借助 SAS 实现典型相关分析,并对输出结果做出解释。

## 1 基本概念

### 1.1 典型变量

典型变量是指从一组变量中提炼出来的一个综合变量,它是无法直接观测的变量,因此,它是一

个隐变量或潜变量。若一组变量中包含  $m$  个定量变量,最多可提取  $m$  个典型变量。这  $m$  个典型变量包含了  $m$  个原变量的全部信息,但它们包含的信息量是逐渐减少的,且它们彼此之间是互相独立的。

### 1.2 典型相关系数

设有两组定量变量,第一组包含  $p$  个定量变量,第二组包含  $q$  个定量变量,不妨设  $p \leq q$ 。采用主成分分析法分别从两组中提取前  $p$  个主成分,它们依次为第一、第二、 $\dots$ 、第  $p$  个主成分,将两组中的第  $i$  个主成分配成一对 ( $i=1, 2, \dots, p$ ),可形成  $p$  个典型变量对。基于 Pearson 相关系数计算公式,可得出每对典型变量之间的相关系数,称此相关系数为典型相关系数。于是,可计算出  $p$  个典型相关系数。

### 1.3 典型相关分析

所谓典型相关分析,实际上就是对两组定量变量进行相关分析,其主要任务包括以下3个方面:其一,呈现出由原变量线性组合而成的各对典型变量;其二,计算并检验各个典型相关系数;其三,结合基本常识和专业知 识,基于各对典型变量表达式中的系数及其正负号,给出解释,并做出统计结论和专业结论。

### 1.4 典型变量个数的确定

在对结果进行分析时,如果第一个典型相关系数经检验无统计学意义,则可认为两组变量之间的相关性无统计学意义,不适合进行典型相关分析。在实际应用中,究竟取几对典型相关变量,除了以典型相关系数有无统计学意义作为参考之外,还需结合典型相关变量对的专业意义来考虑,若有 $k(k>1)$ 个典型相关系数具有统计学意义,就保留 $k$ 个典型变量对。

### 1.5 典型变量的性质

原始变量间、典型变量间、原始变量与典型变量间,存在较复杂的关系,具体地说,可概括成以下3条性质<sup>[1-2]</sup>:①第一组各典型变量之间互不相关,第二组各典型变量之间互不相关,第一组各典型变量与第二组各典型变量两两之间互不相关;②原始变量与典型变量之间存在相关性,具体地说,第一组原始变量与第一组和第二组中典型变量之间存在相关性,第二组原始变量与第一组和第二组中典型变量之间存在相关性;③线性变换不改变相关性,即对两组中的典型变量都做线性变换,再求经线性变换后的新典型变量之间的相关系数,其数值与未经过线性变换的原典型变量之间的相关系数相等。

## 2 计算方法

典型相关分析中的计算主要涉及两方面的内容,即典型相关变量表达式中系数的估计和典型相关系数的假设检验<sup>[3-4]</sup>。

### 2.1 典型相关变量中系数的估计

设 $p \leq q, p+q$ 维随机向量的协方差矩阵 $\Sigma > 0$ ,也就是该矩阵为正定矩阵,可表示如下,见式(1)。

$$\text{Cov} \begin{pmatrix} X \\ Y \end{pmatrix} = \Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix} \quad (1)$$

式(1)中, $\Sigma_{11}$ 为第一类变量的协方差矩阵, $\Sigma_{22}$ 为第二类变量的协方差矩阵, $\Sigma_{12} = \Sigma'_{21}$ 是第一类变量与第二类变量之间的协方差矩阵。当 $\Sigma > 0$ 时, $\Sigma_{11}$ 和 $\Sigma_{22}$ 也都是正定矩阵。

典型相关变量 $U$ 和 $V$ 的方差见式(2)。

$$\begin{cases} \text{Var}(a'X) = a'\Sigma_{11}a = 1 \\ \text{Var}(b'Y) = b'\Sigma_{22}b = 1 \end{cases} \quad (2)$$

求第一对典型相关变量就等价于求 $a=(a_1, a_2, \dots, a_p)'$ 和 $b=(b_1, b_2, \dots, b_q)'$ ,使得在 $U$ 和 $V$ 方差为1的条件下,相关系数取得最大值。相关系数的计算见式(3)。

$$\rho(a'X, b'Y) = a'\Sigma_{12}b \quad (3)$$

这是条件极值问题,应用拉格朗日乘法,见式(4)。

$$F(a, b) = a'\Sigma_{12}b - \frac{\lambda_1}{2}(a'\Sigma_{11}a - 1) - \frac{\lambda_2}{2}(b'\Sigma_{22}b - 1) \quad (4)$$

式(4)中, $\lambda_1$ 和 $\lambda_2$ 为拉格朗日乘子,对上式分别关于 $a, b$ 求偏导数并令其为0,见式(5)。

$$\begin{cases} \frac{\partial F}{\partial a} = \Sigma_{12}b - \lambda_1 \Sigma_{11}a = 0 \\ \frac{\partial F}{\partial b} = \Sigma_{21}a - \lambda_2 \Sigma_{22}b = 0 \end{cases} \quad (5)$$

再分别用 $a'$ 和 $b'$ 左乘式(5),得式(6)。

$$\lambda_1 = \lambda_2 = a'\Sigma_{12}b = \rho(U, V) \quad (6)$$

式(6)中, $\lambda_1$ 和 $\lambda_2$ 是相等的,并且都等于 $U$ 与 $V$ 的相关系数,将它们统一用 $\lambda$ 表示,则式(5)等价于式(7)。

$$\begin{cases} -\lambda \Sigma_{11}a + \Sigma_{12}b = 0 \\ \Sigma_{21}a - \lambda \Sigma_{22}b = 0 \end{cases} \quad (7)$$

上式有非零解的充要条件是系数矩阵的行列式为0,见式(8)。

$$\begin{vmatrix} -\lambda \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & -\lambda \Sigma_{22} \end{vmatrix} = 0 \quad (8)$$

这是一个关于 $\lambda$ 的 $p+q$ 次方程,从中解得 $\lambda$ ,把求得的最大的 $\lambda$ 代入式(7),再求得 $a$ 和 $b$ (满足 $a'\Sigma_{11}a=1, b'\Sigma_{22}b=1$ ),从而可以得出第一对典型相关变量,而最大的 $\lambda$ 即为第一典型相关系数。该方程的根均不大于1,它的 $p$ 个正根依次为 $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_p > 0$ ,如果要计算第 $i$ 对典型相关变量,只需将 $\lambda_i$ 代入式(7)解得相应的典型系数,便可得到 $U_i$ 和 $V_i$ 的表达式,而 $\lambda_i$ 即为第 $i$ 个典型相关系数。

具体计算时,由于矩阵  $\sum_{11}^{-1} \sum_{12} \sum_{22}^{-1} \sum_{21}$  和  $\sum_{22}^{-1} \sum_{21} \sum_{11}^{-1} \sum_{12}$  有相同的非零特征根,依次都是  $\lambda_1^2 \geq \lambda_2^2 \geq \dots \geq \lambda_p^2 > 0$ ,  $a$  和  $b$  分别是这两个矩阵对应于某一特征根的特征向量,所以求解典型相关系数及典型相关变量就可以归结为求矩阵的特征值和特征向量。

前文假定了协方差矩阵  $\sum > 0$ ,如若不然,则  $X_i$  之间或  $Y_j$  之间,也就是类内变量之间存在线性关系,一个或多个变量可以被其他变量线性表示,自然可以将其删去。一般情况下,协方差矩阵为非负定矩阵,此时非零典型相关系数的个数  $m$  等于矩阵  $\sum_{12}$  的秩,即存在  $rank(\sum_{12}) = rank(\sum_{11}^{-1} \sum_{12} \sum_{22}^{-1} \sum_{21}) = m$ ,在求解时可以采用广义逆矩阵。

从相关矩阵出发,也可以计算典型相关系数和典型相关变量,将原始数据进行标准化后,相关矩阵即为相应的协方差矩阵。具体的计算步骤与前述相同,只需用相关矩阵代替协方差矩阵即可。由于典型相关系数  $\rho(U_i, V_i)$  在标准化变换下具有不变性,所以无论采用哪种方法计算,典型相关系数均相同。记  $D_1 = [\text{diag}(\sum_{11})]^{-1/2}$ ,  $D_2 = [\text{diag}(\sum_{22})]^{-1/2}$ ,  $\text{diag}(\sum_{11})$  表示由  $\sum_{11}$  的对角元素组成的对角阵,则得到的典型系数  $a_i^*$ 、 $b_i^*$  与原典型系数  $a_i$ 、 $b_i$  之间的关系如下,见式(9)。

$$a_i = D_1 a_i^*, b_i = D_2 b_i^*, i=1, 2, \dots, P \quad (9)$$

由此则可以得到原始变量的典型相关变量,且原始变量的典型相关变量与根据相关矩阵计算得到的典型相关变量之间只相差一个常数。

### 2.2 典型相关系数的假设检验

当两类变量  $X$  和  $Y$  不相关,即  $c(X, Y) = \sum_{12} = 0$  时,所有的典型相关系数必为零,做典型相关分析毫无意义,所以首先应对  $\sum_{12} = 0$  进行检验,原假设和备择假设分别为  $H_0: \sum_{12} = 0; H_1: \sum_{12} \neq 0$ ,似然比统计量见式(10)。

$$\Lambda_1 = \frac{|S|}{|S_{11}| |S_{22}|} = \prod_{i=1}^p (1 - \hat{\lambda}_i^2) \quad (10)$$

式(10)中,  $|S|$  为样本协方差矩阵的行列式,  $\hat{\lambda}_i$  为样本典型相关系数。由似然比统计量可导出如下统计量,见式(11)。

$$Q_1 = -[n - 1 - \frac{1}{2}(p + q + 1)] \ln \Lambda_1 \quad (11)$$

式(11)中,  $Q_1$  近似服从自由度为  $p \times q$  的  $\chi^2$  分布。对于  $\sum_{12} = 0$  的检验实际上等价于检验  $H_0:$

$\lambda_1 = \lambda_2 = \dots = \lambda_p = 0$ ,如果拒绝  $H_0$ ,则可以得出至少第一典型相关系数不为零,此时可以对其余典型相关系数继续进行检验,原假设及备择假设分别为  $H_0: \lambda_k = \lambda_{k+1} = \dots = \lambda_p = 0 (k=2, 3, \dots, p); H_1: \lambda_k \neq 0$ ,检验统计量见式(12)。

$$Q_k = -[n - k - \frac{1}{2}(p + q + 1)] \ln \Lambda_k,$$

$$\Lambda_k = \prod_{i=k}^p (1 - \hat{\lambda}_i^2), k=2, 3, \dots, p \quad (12)$$

在式(12)中,  $Q_k$  近似服从自由度为  $(p-k+1) \times (q-k+1)$  的  $\chi^2$  分布,如果  $Q_k \geq \chi_{\alpha, (p-k+1) \times (q-k+1)}^2$ ,则拒绝  $H_0$ ,说明  $\lambda_k$  与 0 之间差异有统计学意义。从  $k=2$  开始逐个检验,直到某个  $H_0$  被接受为止,说明自此以后的典型相关系数均为 0。

## 3 实例与 SAS 实现

### 3.1 问题与数据结构

#### 3.1.1 两个实际问题及数据

【例1】某研究者测量 15 名受试者的身体形态以及健康情况的指标,资料见表 1<sup>[2]</sup>。第一类是身体形态变量,有年龄、体重、胸围和日抽烟量;第二类是健康状况变量,有脉搏、收缩压和舒张压。要求选择合适的统计分析方法考查身体形态与健康状况这两类变量之间的相关性。

【例2】以 284 家公司为样本,研究人力资源管理实践与组织绩效之间的关系。人力资源管理实践变量包括人力资源计划( $X_1$ )、正式选拔机制( $X_2$ )、系统培训( $X_3$ )、绩效报酬( $X_4$ )以及内部劳动力市场( $X_5$ );组织绩效变量包括销售收入( $Y_1$ )、净利润( $Y_2$ )、保留必要的员工( $Y_3$ )、招聘到新员工( $Y_4$ )和员工士气( $Y_5$ )。资料见表 2<sup>[2]</sup>。试选择合适的统计分析方法分析人力资源管理实践与组织绩效之间的相关性。

#### 3.1.2 对数据结构的分析

在例 1 中,7 个定量变量分属于两类,第一类是身体形态变量,包含 4 个变量;第二类是健康状况变量,包含 3 个变量。故这是一个单组设计 7 元定量资料。

在例 2 中,10 个定量变量分属于两类,第一类是人力资源管理实践变量,包含 5 个变量;第二类是组织绩效变量,包含 5 个变量。故这是一个单组设计 10 元定量资料。

表 1 受试者身体素质定量指标的测定结果

Table 1 Measurement results of quantitative indicators of physical fitness among subjects

编号	年龄(X <sub>1</sub> )	体重(X <sub>2</sub> )	抽烟量(X <sub>3</sub> )	胸围(X <sub>4</sub> )	脉搏(Y <sub>1</sub> )	收缩压(Y <sub>2</sub> )	舒张压(Y <sub>3</sub> )
1	25	125	30	83.5	70	130	85
2	26	131	25	82.9	72	135	80
3	28	128	35	88.1	75	140	90
4	29	126	40	88.4	78	140	92
5	27	126	45	80.6	73	138	85
6	32	118	20	88.4	70	130	80
7	31	120	18	87.8	68	135	75
8	34	124	25	84.6	70	135	75
9	36	128	25	88.0	75	140	80
10	38	124	23	85.6	72	145	86
11	41	135	40	86.3	76	148	88
12	46	143	45	84.8	80	145	90
13	47	141	48	87.9	82	148	92
14	48	139	50	81.6	85	150	95
15	45	140	55	88.0	88	160	95

表 2 人力资源管理实践变量与组织绩效变量两两之间的相关系数矩阵

Table 2 Correlation coefficient matrix between human resource management practice variables

变量	X <sub>1</sub>	X <sub>2</sub>	X <sub>3</sub>	X <sub>4</sub>	X <sub>5</sub>	Y <sub>1</sub>	Y <sub>2</sub>	Y <sub>3</sub>	Y <sub>4</sub>	Y <sub>5</sub>
X <sub>1</sub>	1									
X <sub>2</sub>	0.32	1								
X <sub>3</sub>	0.56	0.32	1							
X <sub>4</sub>	0.43	0.36	0.38	1						
X <sub>5</sub>	0.51	0.44	0.63	0.51	1					
Y <sub>1</sub>	0.07	0.02	0.03	0.09	0.08	1				
Y <sub>2</sub>	0.10	0.09	0.04	0.07	0.16	0.74	1			
Y <sub>3</sub>	0.19	0.10	0.02	0.21	0.17	0.28	0.27	1		
Y <sub>4</sub>	0.10	0.20	0.05	0.15	0.20	0.11	0.21	0.30	1	
Y <sub>5</sub>	0.10	0.03	-0.01	0.09	0.03	0.33	0.26	0.51	0.38	1

3.1.3 创建 SAS 数据集

分析例 1 资料, 设所需要的 SAS 数据步程序如下:

```
data a1;
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;
```

分析例 2 资料, 设所需要的 SAS 数据步程序如下:

```
data a2(type=corr);
infile cards missover;
_type_='corr';
input _name_ $ X1-X5 Y1-Y5;
if _n_=1 then _type_='n'; else _type_='corr';
cards;
n 284
X1 1
X2 0.32 1
X3 0.56 0.32 1
```

```

X4 0.43 0.36 0.38 1
X5 0.51 0.44 0.63 0.51 1
Y1 0.07 0.02 0.03 0.09 0.08 1
Y2 0.10 0.09 0.04 0.07 0.16 0.74 1
Y3 0.19 0.10 0.02 0.21 0.17 0.28 0.27 1
Y4 0.10 0.20 0.05 0.15 0.20 0.11 0.21
0.30 1
Y5 0.10 0.03 -0.01 0.09 0.03 0.33 0.26
0.51 0.38 1
;
run;

```

### 3.2 用 SAS 实现统计分析

#### 3.2.1 分析例 1 中的资料

设所需要的 SAS 程序如下<sup>[5-6]</sup>:

```

proc cancorr data=a1;
var X1-X4;
with Y1-Y3;
run;

```

【SAS 输出结果及解释】第一对典型变量之间的典型相关系数为 0.957 ( $P=0.005$ ); 第二对典型变量之间的典型相关系数为 0.582 ( $P=0.618$ ); 第三对典型变量之间的典型相关系数为 0.180 ( $P=0.848$ )。由此可知, 只需要考虑第一对典型变量 ( $V_1, W_1$ ) 即可。

基于第一组中原始变量线性表达该组中第一个典型变量  $V_1$  的表达式, 见式(13)。

$$V_1=0.031X_1+0.019X_2+0.058X_3+0.071X_4 \quad (13)$$

基于第二组中原始变量线性表达该组中第一个典型变量  $W_1$  的表达式, 见式(14)。

$$W_1=0.121Y_1+0.020Y_2+0.021Y_3 \quad (14)$$

基于第一组中标准化变量线性表达该组中第一个典型变量  $V_1$  的表达式, 见式(15)。

$$V_1=0.256X_1+0.151X_2+0.694X_3+0.189X_4 \quad (15)$$

基于第二组中标准化变量线性表达该组中第一个典型变量  $W_1$  的表达式, 见式(16)。

$$W_1=0.721Y_1+0.171Y_2+0.142Y_3 \quad (16)$$

将式(15)、式(16)中的系数与式(13)、式(14)中的系数进行比较可知, 经标准化变换后的系数变大了, 故采用式(15)和式(16)所呈现的第一对典型变量 ( $V_1, W_1$ ) 来下专业结论即可。

结论: 第一组中的 4 个变量与第二组中的 3 个变量之间的整体相关性, 主要体现在  $X_3$  与  $Y_1$  之间

的相关性上, 因为它们在各自表达式中的系数都是最大的。相对于同一个表达式中的其他系数而言, 它们在数量上几乎比其他系数之和还要大。结合例 1 的实际情况来看, 脉搏的快慢与抽烟量关系密切。若求出它们之间的简单相关系数为 0.886,  $P<0.001$ 。也就是说, 抽烟量越多的人, 其脉搏数就越大。

#### 3.2.2 分析例 2 中的资料

设所需要的 SAS 程序如下<sup>[5-6]</sup>:

```

proc cancorr data=a2;
var X1-X5;
with Y1-Y5;
run;

```

【SAS 输出结果及解释】第一对典型变量之间的典型相关系数为 0.351 ( $P=0.001$ ); 第二对典型变量之间的典型相关系数为 0.203 ( $P=0.445$ ); 第三对典型变量之间的典型相关系数为 0.107 ( $P=0.877$ ); 第四对典型变量之间的典型相关系数为 0.067 ( $P=0.862$ ); 第五对典型变量之间的典型相关系数为 0.012 ( $P=0.843$ )。由此可知, 只需要考虑第一对典型变量 ( $V_1, W_1$ ) 即可。

参照本文例 1 中的 4 个表达式 [见式(13)、式(14)、式(15)、式(16)], 基于例 2 输出的结果, 可以写出 4 个表达式, 为节省篇幅, 此处从略。

结论: 第一对典型变量 ( $V_1, W_1$ ) 之间的相关性主要取决于 ( $X_3, X_5$ ) 与 ( $Y_3, Y_4$ ) 之间的相关性。然而,  $X_5$  与 ( $Y_3, Y_4$ ) 之间呈正相关, 而  $X_3$  与 ( $Y_3, Y_4$ ) 之间呈负相关。欲对所得出的上述结果做出合理解释, 需要人力资源管理方面的专业知识, 因篇幅所限, 此处从略。

## 4 讨论与小结

### 4.1 讨论

典型相关分析, 是研究两组变量之间相关性的一种多元统计分析方法。事实上, 典型相关分析是由若干个彼此互相独立的典型变量对之间的相关分析组成的。典型变量对中的每个典型变量是对其在组中全部变量信息的提炼和集成, 在本质上, 典型变量就是主成分。在多数实际问题中, 通常只需要把注意力集中在对第一对典型变量进行研究和解释上即可。

典型相关系数的平方表示两组典型变量间享有的共同变异的百分比, 可进一步分解为各自的解

释能力。将“解释能力”乘以典型相关系数的平方,它表示一组变量的方差被另一组典型变量解释的平均比例,该比例也被称为典型冗余。因篇幅所限,此处从略,可参阅文献[3,7]。

每组究竟应保留几个典型变量,通常依据经验来确定,但文献[8]中给出了通过假设检验来确定的方法。文献[1]中给出典型相关分析的几何解释,还给出了基于典型变量计算典型得分和进行预测的方法。

## 4.2 小结

本文介绍了与定量变量典型相关分析有关的基本概念、计算方法、两个实例及其SAS实现。基本概念包括典型变量、典型相关系数、典型相关分析、典型变量个数的确定和典型变量的性质;计算方法涉及典型相关变量中系数的估计和典型相关系数的假设检验;两个实例的资料分别是“受试者的身体形态及健康指标的测定结果”和“人力资源管理实践变量与组织绩效变量的测定结果”。借助SAS对两个实例中的定量资料进行了典型相关分析,并对输出结果做出了合理解释。

## 参考文献

- [1] 张润楚. 多元统计分析[M]. 北京: 科学出版社, 2006: 271-287.  
Zhang RC. Multivariate statistical analysis [M]. Beijing: Science Press, 2006: 271-287.
- [2] 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: 84-106.  
Hu LP. Problem-oriented statistics: (3) experimental design and multivariate statistical analysis [M]. Beijing: People's Mental Publishing House, 2012: 84-106.
- [3] 高惠璇. 应用多元统计分析[M]. 北京: 北京大学出版社, 2005: 343-368.  
Gao HX. Applied multivariate statistical analysis [M]. Beijing: Peking University Press, 2005: 343-368.
- [4] 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: 539-574.  
Johnson RA, Wichern DW. Applied multivariate statistical analysis [M]. 6<sup>th</sup> edition. Beijing: Tsinghua University Press, 2008: 539-574.
- [5] SAS Institute Inc. SAS/STAT®15.1 user's guide[M]. Cary, NC: SAS Institute Inc, 2018: 2063-2092.
- [6] 何晓群. 多元统计分析[M]. 2版. 北京: 中国人民大学出版社, 2008: 255-282.  
He XQ. Multivariate statistical analysis [M]. 2<sup>nd</sup> edition. Beijing: ChinaRenmin University Press, 2008: 255-282.
- [7] 李卫东. 应用多元统计分析[M]. 北京: 北京大学出版社, 2008: 259-273.  
Li WD. Applied multivariate statistical analysis [M]. Beijing: Peking University Press, 2008: 259-273.
- [8] 王静龙. 多元统计分析[M]. 北京: 科学出版社, 2008: 333-346.  
Wang JL. Multivariate statistical analysis [M]. Beijing: Science Press, 2008: 333-346.

(收稿日期:2023-06-05)

(本文编辑:陈霞)