

典型相关分析

Canonical Correlation Analysis

研究多个变量与多个变量之间的相关性

要 点

- 典型相关分析的数学表达方式，约束条件；
- 典型相关系数的数学含义；
- 典型变量的数学含义；
- 典型相关系数的显著性检验；
- 冗余分析；
- 典型相关的应用

第一节 典型相关分析的基本思想

当研究两个变量x与y之间的相关关系时，**相关系数**是最常用的度量。

$$\rho(x, y) = \frac{\text{cov}(x, y)}{\sqrt{s_{xx}s_{yy}}}$$

如何研究**两组**变量之间的相关关系呢？

x_1	y_1	$r_{x_1y_1}$	$r_{x_1y_2}$
x_2	y_2	$r_{x_2y_1}$	$r_{x_2y_2}$
x_3		$r_{x_3y_1}$	$r_{x_3y_2}$

如何进一步确定两组变量在**整体**上的相关程度呢？

通常情况下，为了研究两组变量

$$(x_1, x_2, \dots, x_p) \quad (y_1, y_2, \dots, y_q)$$

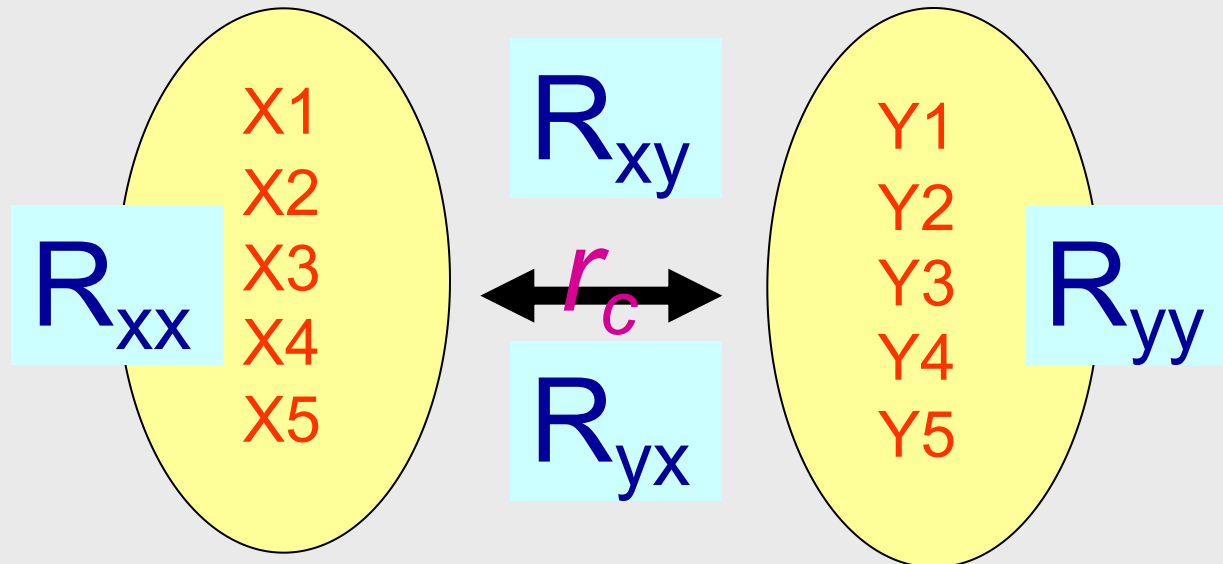
的相关关系，可以用最原始的方法，分别计算两组变量之间的全部相关系数，一共有 pq 个简单相关系数，这样既烦琐又不能抓住问题的本质。

如果能够采用类似于主成分的思想，分别找出两组变量的各自的某个线性组合，讨论线性组合之间的相关关系，则更简捷。

- 典型相关分析（Canonical Correlation）是研究两组变量之间相关关系的一种多元统计方法。它能够揭示两组变量之间的内在联系。

基本概念

- Analyze the relationships between two sets of variables
- Canonical correlation(r_c): Correlation between two composition of variables



- 1936年霍特林（Hotelling）最早就“大学表现”和“入学前成绩”的关系、政府政策变量与经济目标变量的关系等问题进行了研究，提出了典型相关分析技术。
- 之后，Cooley和Hohnes(1971)，Tatsuoka(1971)及Mardia, Kent和Bibby(1979)等人对典型相关分析的应用进行了讨论，Kshirsagar(1972)则从理论上给出了最好的分析。

在解决实际问题中，这种方法有广泛的应用。如居民生活环境与健康状况的关系；考察一些与**财政政策**有关的指标：财政支出总额增长率、财政赤字增长率、税率降低，与**经济发展**的一系列指标如**GDP**增长率、就业增长率、物价上涨率等，来研究**扩张性财政政策**实施后对**宏观经济发展**的影响。这些多变量间的相关性如何分析？

- 典型相关分析的目的在于识别并量化两组变量之间的联系，将两组变量相关关系的分析，转化为一组变量的线性组合与另一组变量线性组合之间的相关关系分析。
- 目前，典型相关分析已被应用于心理学、市场营销等领域。如用于研究个人性格与职业兴趣的关系，市场促销活动与消费者响应之间的关系等问题的分析研究。

利用主成分分析的思想，可以把多个变量与多个变量之间的相关转化为两个变量之间的相关。

主成分

$$\underset{\text{综合变量}}{y_1} = a_{11}x_1 + a_{21}x_2 + \dots + a_{p1}x_p$$

找出系数 (a_1, a_2, a_3) 和 (b_1, b_2) 使得新变量

$$u = a_1x_1 + a_2x_2 + a_3x_3 \quad \text{和} \quad v = b_1y_1 + b_2y_2$$

之间有最大可能的相关系数。

（典型相关系数）

即使 $\rho(u, v) \rightarrow \max$

例 家庭特征与家庭消费之间的关系

为了了解家庭的特征与其消费模式之间的关系。

调查了70个家庭的下面两组变量：

$$\begin{cases} x_1: & \text{每年去餐馆就餐的频率} \\ x_2: & \text{每年外出看电影频率} \end{cases}$$
$$\begin{cases} y_1: & \text{户主的年龄} \\ y_2: & \text{家庭的年收入} \\ y_3: & \text{户主受教育程度} \end{cases}$$

分析两组变量之间的关系。

变量间的相关系数矩阵

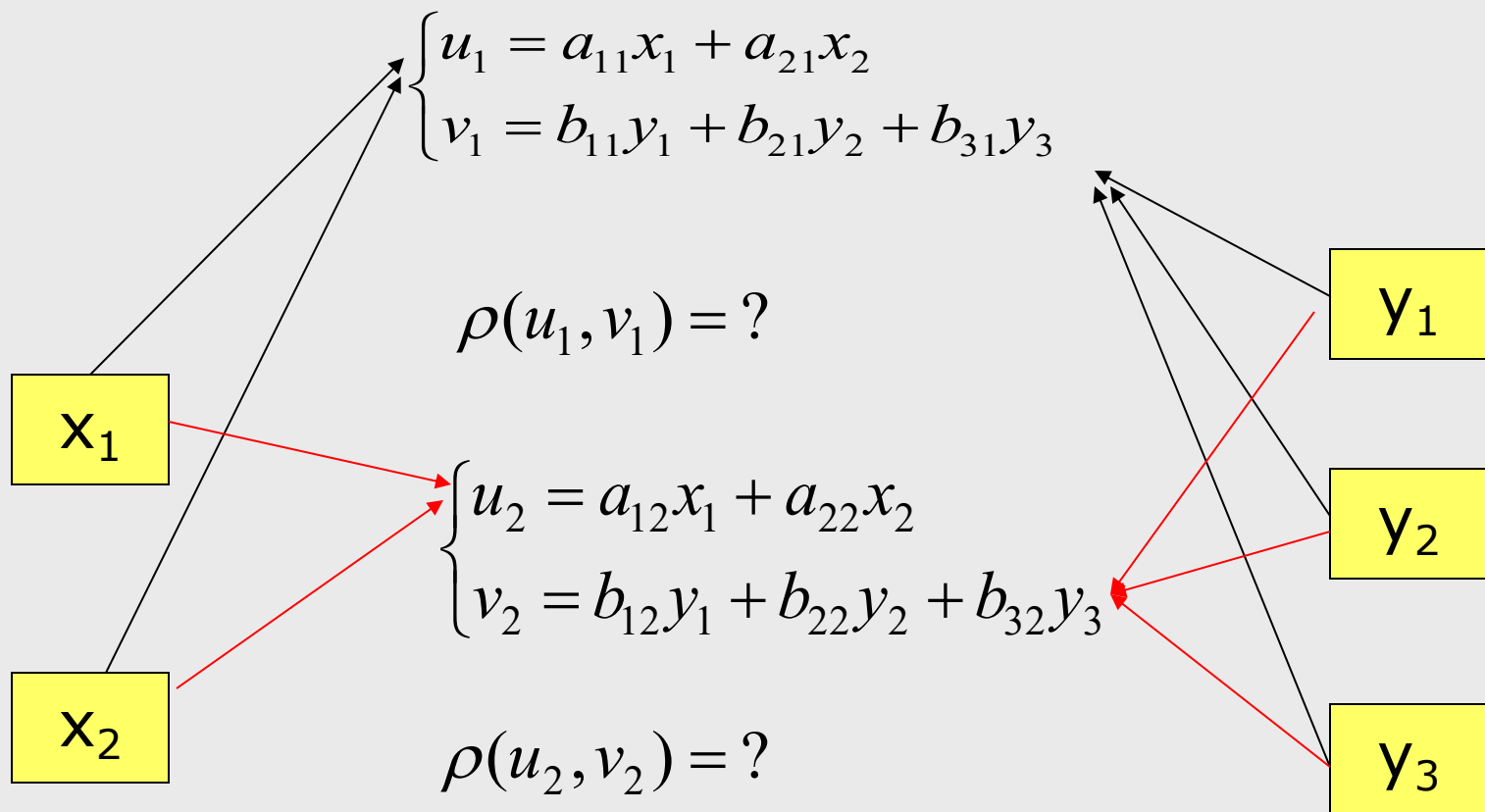
	X1	X2	y1	y2	y3
X1	1.00	0.80	0.26	0.67	0.34
X2	0.80	1.00	0.33	0.59	0.34
y1	0.26	0.33	1.00	0.37	0.21
y2	0.67	0.59	0.37	1.00	0.35
y3	0.34	0.34	0.21	0.35	1.00

R_{11}

R_{12}

R_{21}

R_{22}



典型相关分析的思想：

首先分别在每组变量中找出第一对线性组合，使其具有最大相关性，

$$\begin{cases} u_1 = a_{11}x_1 + a_{21}x_2 + \cdots + a_{p1}x_p \\ v_1 = b_{11}y_1 + b_{21}y_2 + \cdots + b_{q1}y_q \end{cases}$$

然后再在每组变量中找出第二对线性组合，使其分别与本组内的第一线性组合不相关，第二对本身具有次大的相关性。

$$\begin{cases} u_2 = a_{12}x_1 + a_{22}x_2 + \cdots + a_{p\ 2}x_p \\ v_2 = b_{12}y_1 + b_{22}y_2 + \cdots + b_{q\ 2}y_q \end{cases}$$

u_2 和 v_2 与 u_1 和 v_1 相互独立，但 u_2 和 v_2 相关。如此继续下去，直至进行到 r 步，两组变量的相关性被提取完为止。 $r \leq \min(p, q)$ ，可以得到 r 组变量。

二、典型相关的数学描述

(一) 想法

考虑两组变量的向量 $\mathbf{Z} = (x_1, x_2, \dots, x_p, y_1, y_2, \dots, y_q)$

其协方差阵为 $\Sigma = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix}$

$p \qquad q$

其中 Σ_{11} 是第一组变量的协方差矩阵； Σ_{22} 是第二组变量的协方差矩阵； $\Sigma_{12} = \Sigma'_{21}$ 是X和Y的协方差矩阵。

如果我们记两组变量的第一对线性组合为：

$$\mathbf{u}_1 = \mathbf{a}_1' \mathbf{X} \quad \mathbf{v}_1 = \mathbf{b}_1' \mathbf{Y}$$

其中： $\mathbf{a}_1 = (a_{11}, a_{21}, \dots, a_{p1})'$

$$\mathbf{b}_1 = (b_{11}, b_{21}, \dots, b_{q1})'$$

$$Var(u_1) = \mathbf{a}_1' Var(\mathbf{X}) \mathbf{a}_1 = \mathbf{a}_1' \boldsymbol{\Sigma}_{11} \mathbf{a}_1 = 1$$

$$Var(\mathbf{v}_1) = \mathbf{b}_1' Var(\mathbf{Y}) \mathbf{b}_1 = \mathbf{b}_1' \boldsymbol{\Sigma}_{22} \mathbf{b}_1 = 1$$

$$\rho_{\mathbf{u}_1, \mathbf{v}_1} = Cov(u_1, v_1) = \mathbf{a}_1' Cov(\mathbf{X}, \mathbf{Y}) \mathbf{b}_1 = \mathbf{a}_1' \boldsymbol{\Sigma}_{12} \mathbf{b}_1$$

所以，典型相关分析就是求 \mathbf{a}_1 和 \mathbf{b}_1 ，使 ρ_{uv} 达到最大。

(二) 典型相关系数和典型变量的求法

在约束条件:

$$Var(u) = \mathbf{a}'\mathbf{\Sigma}_{11}\mathbf{a} = 1$$

$$Var(v) = \mathbf{b}'\mathbf{\Sigma}_{22}\mathbf{b} = 1$$

下, 求 \mathbf{a}_1 和 \mathbf{b}_1 , 使 ρ_{uv} 达到最大。令

根据数学分析中条件极值的求法，引入Lagrange乘数，求极值问题，则可以转化为求

$$\phi(a_1, b_1) = a_1' \Sigma_{12} b_1 - \frac{\lambda}{2} (a_1' \Sigma_{11} a_1 - 1) - \frac{\nu}{2} (b_1' \Sigma_{22} b_1 - 1) \quad (1)$$

的极大值，其中 λ 和 ν 是 Lagrange乘数。

$$\begin{cases} \frac{\partial \phi}{\partial a_1} = \Sigma_{12} b_1 - \lambda \Sigma_{11} a_1 = 0 \\ \frac{\partial \phi}{\partial b_1} = \Sigma_{21} a_1 - v \Sigma_{22} b_1 = 0 \end{cases} \quad (2)$$

$$\begin{cases} \Sigma_{12} \mathbf{b}_1 - \lambda \Sigma_{11} \mathbf{a}_1 = \mathbf{0} \\ \Sigma_{21} \mathbf{a}_1 - v \Sigma_{22} \mathbf{b}_1 = \mathbf{0} \end{cases} \quad (3)$$

将上面的3式分别左乘 \mathbf{a}_1' 和 \mathbf{b}_1'

$$\begin{cases} \mathbf{a}_1' \Sigma_{12} \mathbf{b}_1 - \lambda \mathbf{a}_1' \Sigma_{11} \mathbf{a}_1 = 0 \\ \mathbf{b}_1' \Sigma_{21} \mathbf{a}_1 - v \mathbf{b}_1' \Sigma_{22} \mathbf{b}_1 = 0 \end{cases}$$

$$\begin{cases} \mathbf{a}_1' \Sigma_{12} \mathbf{b}_1 = \lambda \\ \mathbf{b}_1' \Sigma_{21} \mathbf{a}_1 = v \end{cases}$$

$$\begin{cases} \Sigma_{12} \mathbf{b}_1 - \lambda \Sigma_{11} \mathbf{a}_1 = \mathbf{0} \\ \Sigma_{21} \mathbf{a}_1 - \lambda \Sigma_{22} \mathbf{b}_1 = \mathbf{0} \end{cases} \quad (3)$$

$$\lambda = v = \mathbf{a}_1' \Sigma_{12} \mathbf{b}_1 = \rho(u_1, v_1)$$

由 (3) 的第一式, 得 $\Sigma_{12} b_1 = \lambda \Sigma_{11} \mathbf{a}_1$

$$b_1 = \Sigma_{12}^{-1} \lambda \Sigma_{11} \mathbf{a}_1$$

问题是, Σ_{12}^{-1} 是否存在

$$\begin{cases} \Sigma_{12}\mathbf{b}_1 - \lambda\Sigma_{11}\mathbf{a}_1 = \mathbf{0} \\ \Sigma_{21}\mathbf{a}_1 - \lambda\Sigma_{22}\mathbf{b}_1 = \mathbf{0} \end{cases} \quad (3)$$

$$\Sigma_{12}\mathbf{b}_1 = \lambda\Sigma_{11}\mathbf{a}_1$$

将 $\Sigma_{12}\Sigma_{22}^{-1}$ 左乘 (3) 的第二式，得

$$\Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}\mathbf{a}_1 - \lambda\Sigma_{12}\Sigma_{22}^{-1}\Sigma_{22}\mathbf{b}_1 = \mathbf{0}$$

$$\Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}\mathbf{a}_1 - \lambda\Sigma_{12}\mathbf{b}_1 = \mathbf{0}$$

并将 (3) 的第一式代入，得

$$\Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}\mathbf{a}_1 - \lambda^2\Sigma_{11}\mathbf{a}_1 = \mathbf{0}$$

$$\Sigma_{11}^{-1}\Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}\mathbf{a}_1 - \lambda^2\mathbf{a}_1 = \mathbf{0}$$

$\Sigma_{11}^{-1}\Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}$ 的特征根
是 λ^2 ，相应的特征向
量为 \mathbf{a}_1

$$\begin{cases} \Sigma_{12}\mathbf{b}_1 - \lambda\Sigma_{11}\mathbf{a}_1 = \mathbf{0} \\ \Sigma_{21}\mathbf{a}_1 - \lambda\Sigma_{22}\mathbf{b}_1 = \mathbf{0} \end{cases} \quad (3)$$

将 $\Sigma_{21}\Sigma_{11}^{-1}$ 左乘 (3) 的第一式，并将第二式代入，得

$$\Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12}\mathbf{b}_1 - \lambda\Sigma_{21}\mathbf{a}_1 = \mathbf{0}$$

$$\Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12}\mathbf{b}_1 - \lambda^2\Sigma_{22}\mathbf{b}_1 = \mathbf{0}$$

$$\Sigma_{22}^{-1}\Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12}\mathbf{b}_1 - \lambda^2\mathbf{b}_1 = \mathbf{0}$$

$\Sigma_{22}^{-1}\Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12}$ 的特征根
是 λ^2 ，相应的特征向
量为 \mathbf{b}_1

令

$$\mathbf{M}_1 = \Sigma_{11}^{-1}\Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}$$

$$\mathbf{M}_2 = \Sigma_{22}^{-1}\Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12}$$

$$\begin{cases} \mathbf{M}_1\mathbf{a} = \lambda^2\mathbf{a} \\ \mathbf{M}_2\mathbf{b} = \lambda^2\mathbf{b} \end{cases}$$

结论: λ^2 既是 M_1 又是 M_2 的特征根, a_1 和 b_1 是相应于 M_1 和 M_2 的特征向量。

至此, 典型相关分析转化为求 M_1 和 M_2 特征根和特征向量的问题。

第一对典型变量提取了原始变量 X 与 Y 之间相关的主要部分, 如果这部分还不能足以解释原始变量, 可以在剩余的相关中再求出第二对典型变量和他们的典型相关系数。

在剩余的相关中再求出第二对典型变量和他们的典型相关系数。设第二对典型变量为：

$$u_2 = \mathbf{a}'_2 \mathbf{x} \quad \mathbf{v}_2 = \mathbf{b}'_2 \mathbf{y}$$

在约束条件： $Var(u_2) = \mathbf{a}'_2 \boldsymbol{\Sigma}_{11} \mathbf{a}_2 = 1$

$$Var(v_2) = \mathbf{b}'_2 \boldsymbol{\Sigma}_{22} \mathbf{b}_2 = 1$$

$$\text{cov}(u_1, u_2) = \text{cov}(\mathbf{a}'_1 \mathbf{x}, \mathbf{a}'_2 \mathbf{x}) = \mathbf{a}'_1 \boldsymbol{\Sigma}_{11} \mathbf{a}_2 = 0$$

$$\text{cov}(v_1, v_2) = \text{cov}(\mathbf{b}'_1 \mathbf{y}, \mathbf{b}'_2 \mathbf{y}) = \mathbf{b}'_1 \boldsymbol{\Sigma}_{22} \mathbf{b}_2 = 0$$

求使 $\text{cov}(u_2, v_2) = \mathbf{a}'_2 \boldsymbol{\Sigma}_{12} \mathbf{b}_2$ 达到最大的 a_2 和 b_2 。

例 家庭特征与家庭消费之间的关系

为了了解家庭的特征与其消费模式之间的关系。

调查了70个家庭的下面两组变量：

$\begin{cases} x_1: & \text{每年去餐馆就餐的频率} \\ x_2: & \text{每年外出看电影频率} \end{cases}$

$\begin{cases} y_1: & \text{户主的年龄} \\ y_2: & \text{家庭的年收入} \\ y_3: & \text{户主受教育程度} \end{cases}$

分析两组变量之间的关系。

变量间的相关系数矩阵

	X1	X2	y1	y2	y3
X1	1.00	0.80	0.26	0.67	0.34
X2	0.80	1.00	0.33	0.59	0.34
y1	0.26	0.33	1.00	0.37	0.21
y2	0.67	0.59	0.37	1.00	0.35
y3	0.34	0.34	0.21	0.35	1.00

典型相关分析				
	典型相 关系数	调整典型 关系数	近似方差	典型相关系 数的平方
1	0.687948	0.687848	0.005268	0.473272
2	0.186865	0.186638	0.009651	0.034919

X组典型变量的系数		
	U1	U2
X1（就餐）	0.7689	-1.4787
X2（电影）	0.2721	1.6443
Y组典型变量的系数		
	V1	V2
Y1（年龄）	0.0491	1.0003
Y2（收入）	0.8975	-0.5837
Y3（文化）	0.1900	0.2956

$$u_1 = 0.7689x_1 + 0.2721x_2 \qquad v_1 = 0.0491y_1 + 0.8975y_2 + 0.1900y_3$$

$$u_2 = -1.4787x_1 + 1.6443x_2 \qquad v_2 = 1.0003y_1 - 0.5837y_2 + 0.2956y_3$$

三、典型变量的性质

1、同一组变量的典型变量之间互不相关

$$u_k = \mathbf{a}'_k \mathbf{x} \quad v_k = \mathbf{b}'_k \mathbf{y} \quad k, l = 1, 2, \dots, r; k \neq l$$

因为特征向量之间是正交的。故

X组的典型变量之间是相互独立的：

$$\text{cov}(u_k, u_l) = \text{cov}(a'_k X, a'_l X) = a'_k \Sigma_{11} a_l = 0$$

Y组的典型变量之间是相互独立的：

$$\text{cov}(v_k, v_l) = \text{cov}(b'_k Y, b'_l Y) = b'_k \Sigma_{11} b_l = 0$$

2、不同组变量的典型变量之间的相关性

不同组内一对典型变量之间的相关系数为：

$$\begin{aligned}\text{cov}(u_i, v_j) &= \text{cov}(\mathbf{a}_i' \mathbf{x}, \mathbf{b}_j' \mathbf{y}) \\ &= \mathbf{a}_i' \text{cov}(\mathbf{x}, \mathbf{y}) \mathbf{b}_j' = \mathbf{a}_i' \boldsymbol{\Sigma}_{12} \mathbf{b}_j \\ &= \begin{cases} \lambda_i, i = j \\ 0, i \neq j \end{cases}\end{aligned}$$

同对相关系数为 λ_i ，不同对则为零。

3、原始变量与典型变量之间的相关系数 (典型载荷分析)

原始变量相关系数矩阵

$$R = \begin{bmatrix} R_{11} & R_{12} \\ R_{21} & R_{22} \end{bmatrix}$$

X典型变量系数矩阵

$$\mathbf{A} = [\mathbf{a}_1 \quad \mathbf{a}_2 \quad \cdots \quad \mathbf{a}_r]_{p \times r} = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1r} \\ a_{21} & a_{22} & \cdots & a_{2r} \\ \vdots & \vdots & & \vdots \\ a_{p1} & a_{p2} & \cdots & a_{pr} \end{bmatrix}$$

y典型变量系数矩阵

$$\mathbf{B} = [\mathbf{b}_1 \quad \mathbf{b}_2 \quad \cdots \quad \mathbf{b}_r]_{q \times r} = \begin{bmatrix} b_{11} & b_{12} & \cdots & b_{1r} \\ b_{21} & b_{22} & \cdots & b_{2r} \\ \vdots & \vdots & & \vdots \\ b_{q1} & b_{q2} & \cdots & b_{qr} \end{bmatrix}$$

$$\text{cov}(x_i, u_j) = \text{cov}(x_i, a_{1j}x_1 + a_{2j}x_2 + \cdots + a_{pj}x_p)$$

$$= \text{cov}(x_i, a_{1j}x_1) + \text{cov}(x_i, a_{2j}x_2) + \cdots + \text{cov}(x_i, a_{pj}x_p)$$

$$= \sum_{k=1}^p a_{kj} \sigma_{x_i, x_k}$$

$$\rho(x_i, u_j) = \sum_{k=1}^p a_{kj} \sigma_{x_i, x_k} / \sqrt{\sigma_{x_i x_i}}$$

$$\text{cov}(x_i, v_j) = \text{cov}(x_i, b_{1j}y_1 + b_{2j}y_2 + \cdots + b_{pj}y_p)$$

$$= \text{cov}(x_i, b_{1j}y_1) + \text{cov}(x_i, b_{2j}y_2) + \cdots + \text{cov}(x_i, b_{pj}y_p)$$

$$= \sum_{k=1}^q b_{kj} \sigma_{x_i, y_k}$$

$$\rho(x_i, v_j) = \sum_{k=1}^q b_{kj} \sigma_{x_i, y_k} / \sqrt{\sigma_{x_i x_i}}$$

$$\begin{aligned}
\text{cov}(y_i, u_j) &= \text{cov}(y_i, a_{1j}x_1 + a_{2j}x_2 + \cdots + a_{pj}x_p) \\
&= \text{cov}(y_i, a_{1j}x_1) + \text{cov}(y_i, a_{2j}x_2) + \cdots + \text{cov}(y_i, a_{pj}x_p) \\
&= \sum_{k=1}^p a_{kj} \sigma_{y_i, x_k}
\end{aligned}$$

$$\rho(y_i, u_j) = \sum_{k=1}^p a_{kj} \sigma_{y_i, x_k} / \sqrt{\sigma_{y_i y_i}}$$

$$\text{cov}(y_i, v_j) = \text{cov}(y_i, b_{1j}y_1 + b_{2j}y_2 + \cdots + b_{pj}y_p)$$

$$= \text{cov}(y_i, b_{1j}y_1) + \text{cov}(y_i, b_{2j}y_2) + \cdots + \text{cov}(y_i, b_{pj}y_p)$$

$$= \sum_{k=1}^q b_{kj} \sigma_{y_i, y_k}$$

$$\rho(y_i, v_j) = \sum_{k=1}^q b_{kj} \sigma_{y_i, y_k} / \sqrt{\sigma_{y_i y_i}}$$

例 家庭特征与家庭消费之间的关系

为了了解家庭的特征与其消费模式之间的关系。

调查了70个家庭的下面两组变量：

$$\left\{ \begin{array}{l} x_1: \text{每年去餐馆就餐的频率} \\ x_2: \text{每年外出看电影频率} \end{array} \right. \quad \left\{ \begin{array}{l} y_1: \text{户主的年龄} \\ y_2: \text{家庭的年收入} \\ y_3: \text{户主受教育程度} \end{array} \right.$$

分析两组变量之间的关系。

变量间的相关系数矩阵

	x1	x2	y1	y2	y3
x1	1.00	0.80	0.26	0.67	0.34
x2	0.80	1.00	0.33	0.59	0.34
y1	0.26	0.33	1.00	0.37	0.21
y2	0.67	0.59	0.37	1.00	0.35
y3	0.34	0.34	0.21	0.35	1.00

典型相关分析

	典型相 关系数	调整典型 关系数	近似方差	典型相关系 数的平方
1	0.687948	0.687848	0.005268	0.473272
2	0.186865	0.186638	0.009651	0.034919

X组典型变量的系数		
	U1	U2
X1（就餐）	0.7689	-1.4787
X2（电影）	0.2721	1.6443
Y组典型变量的系数		
	V1	V2
Y1（年龄）	0.0491	1.0003
Y2（收入）	0.8975	-0.5837
Y3（文化）	0.1900	0.2956

$$u_1 = 0.7689x_1 + 0.2721x_2$$

$$v_1 = 0.0491y_1 + 0.8975y_2 + 0.1900y_3$$

$$u_2 = -1.4787x_1 + 1.6443x_2$$

$$v_2 = 1.0003y_1 - 0.5837y_2 + 0.2956y_3$$

典型载荷分析

典型变量的结构（相关系数）

	U1	U2
X1	0.9866	-0.1632
X2	0.8872	0.4614
	V1	V2
Y1	0.4211	0.8464
Y2	0.9822	-0.1101
Y3	0.5145	0.3013

典型变量的结构（相关系数）		Cross loadings
	V1	V2
X1	0.6787	-0.0305
X2	0.6104	0.0862
	U1	U2
Y1	0.2897	0.1582
Y2	0.6757	-0.0206
Y3	0.3539	0.0563

两个反映消费的指标与第一对典型变量中 u_1 的相关系数分别为0.9866和0.8872，可以看出 u_1 可以作为消费特性的指标，第一对典型变量中 v_1 与 Y_2 之间的相关系数为0.9822，可见典型变量 v_1 主要代表了了家庭收入， u_1 和 v_1 的相关系数为0.6879，这就说明家庭的消费与家庭的收入之间其关系是很密切的；

第二对典型变量中 u_2 与 x_2 的相关系数为0.4614，可以看出 u_2 可以作为文化消费特性的指标，第二对典型变量中 v_2 与 Y_1 和 Y_3 之间的分别相关系数为0.8464和0.3013，可见典型变量 v_2 主要代表了家庭成员的年龄特征和教育程度， u_2 和 v_2 的相关系数为0.1869，说明文化消费与年龄和受教育程度之间的相关性。

4、各组原始变量被典型变量所解释的方差 (典型冗余分析)

X组原始变量被 u_i 解释的方差比例

$$m_{u_i} = (\rho_{u_i, x_1}^2 + \rho_{u_i, x_2}^2 + \cdots + \rho_{u_i, x_p}^2) / p$$

X组原始变量被 v_i 解释的方差比例

$$m_{v_i} = (\rho_{v_i, x_1}^2 + \rho_{v_i, x_2}^2 + \cdots + \rho_{v_i, x_p}^2) / p$$

y组原始变量被 u_i 解释的方差比例

$$n_{u_i} = (\rho_{u_i, y_1}^2 + \rho_{u_i, y_2}^2 + \cdots + \rho_{u_i, y_q}^2) / q$$

y组原始变量被 v_i 解释的方差比例

$$n_{v_i} = (\rho_{v_i, y_1}^2 + \rho_{v_i, y_2}^2 + \cdots + \rho_{v_i, y_q}^2) / q$$

被典型变量解释的X组原始变量的方差

	被本组的典型变量解释		被对方Y组典型变量解释		
	比例	累计比例	典型相关 系数平方	比例	累计比例
1	0.8803	0.8803	0.4733	0.4166	0.4166
2	0.1197	1.0000	0.0349	0.0042	0.4208

被典型变量解释的Y组原始变量的方差

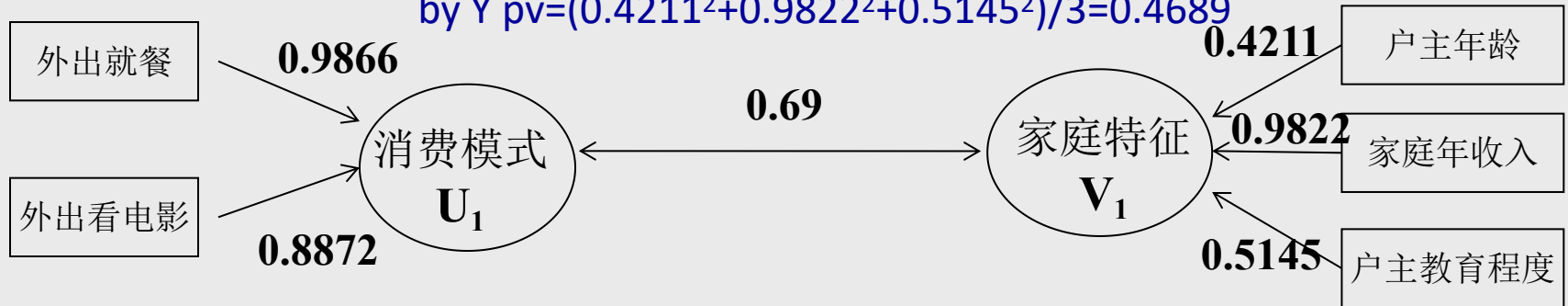
	被本组的典型变量解释		被对方X组典型变量解释		
	比例	累计比例	典型相关 系数平方	比例	累计比例
1	0.4689	0.4689	0.4733	0.2219	0.2219
2	0.2731	0.7420	0.0349	0.0095	0.2315

Path diagram for canonical analysis

Proportion of variance extracted

$$\text{by X } pv = (0.9866^2 + 0.8872^2)/2 = 0.8803$$

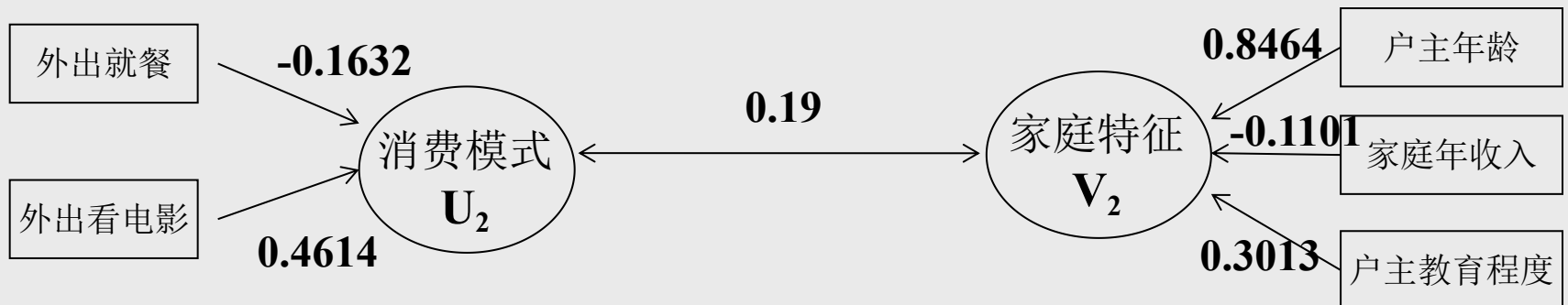
$$\text{by Y } pv = (0.4211^2 + 0.9822^2 + 0.5145^2)/3 = 0.4689$$



Proportion of variance extracted

$$\text{by X } pv = (-0.1632^2 + 0.4614^2)/2 = 0.1197$$

$$\text{by Y } pv = (0.8464^2 + (-0.1101)^2 + 0.3013^2)/3 = 0.2731$$



五、样本典型相关系数

在实际应用中，总体的协方差矩阵常常是未知的，类似于其他的统计分析方法，需要从总体中抽出一个样本，根据样本对总体的协方差或相关系数矩阵进行估计，然后利用估计得到的协方差或相关系数矩阵进行分析。

1、假设有X组和Y组变量，样本容量为n。假设 $(X_1, Y_1), (X_2, Y_2), \dots, (X_n, Y_n)$ ，观测值矩阵为：

$$\begin{bmatrix} x_{11} & \cdots & x_{1p} & y_{11} & \cdots & y_{1q} \\ x_{21} & \cdots & x_{2p} & y_{21} & \cdots & y_{2q} \\ x_{31} & \cdots & x_{3p} & y_{31} & \cdots & y_{3q} \\ x_{41} & \cdots & x_{4p} & y_{41} & \cdots & y_{4q} \\ \vdots & & \vdots & \vdots & & \vdots \\ x_{n1} & \cdots & x_{np} & y_{n1} & \cdots & y_{nq} \end{bmatrix}$$

样本的协方差： $\hat{\Sigma} = \begin{pmatrix} S_{xx} & S_{xy} \\ S_{yx} & S_{yy} \end{pmatrix}$

2、计算特征根和特征向量

$$\text{令: } \hat{M}_1 = (S_{xx}^{-1} S_{xy} S_{yy}^{-1} S_{yx}) \quad \text{令: } \hat{M}_2 = (S_{yy}^{-1} S_{yx} S_{xx}^{-1} S_{xy})$$

求 \mathbf{M}_1 和 \mathbf{M}_2 的特征根 $\lambda_1^2 \geq \lambda_2^2 \geq \cdots \geq \lambda_r^2$ ，对应的特征向量 α_i 和 $\beta_i (i=1,2,\cdots,r)$ 。则特征向量构成典型变量的系数，特征根为典型变量相关系数的平方。

职业满意度典型相关分析

某调查公司从一个大型零售公司随机调查了784人，测量了5个职业特性指标和7个职业满意度变量。讨论两组指标之间是否相联系。

X组：

X1—用户反馈

X2—任务重要性

X3—任务多样性

X4—任务特殊性

X5—自主权

Y组：

Y1—主管满意度

Y2—事业前景满意度

Y3—财政满意度

Y4—工作强度满意度

Y5—公司地位满意度

Y6—工作满意度

Y7—总体满意度

	X1	X2	X3	X4	X5	Y1	Y2	Y3	Y4	Y5	Y6	Y7
X1	1.00	0.49	0.53	0.49	0.51	0.33	0.32	0.20	0.19	0.30	0.37	0.21
X2	0.49	1.00	0.57	0.46	0.53	0.30	0.21	0.16	0.08	0.27	0.35	0.20
X3	0.53	0.57	1.00	0.48	0.57	0.31	0.23	0.14	0.07	0.24	0.37	0.18
X4	0.49	0.46	0.48	1.00	0.57	0.24	0.22	0.12	0.19	0.21	0.29	0.16
X5	0.51	0.53	0.57	0.57	1.00	0.38	0.32	0.17	0.23	0.32	0.36	0.27
Y1	0.33	0.30	0.31	0.24	0.38	1.00	0.43	0.27	0.24	0.34	0.37	0.40
Y2	0.32	0.21	0.23	0.22	0.32	0.43	1.00	0.33	0.26	0.54	0.32	0.58
Y3	0.20	0.16	0.14	0.12	0.17	0.27	0.33	1.00	0.25	0.46	0.29	0.45
Y4	0.19	0.08	0.07	0.19	0.23	0.24	0.26	0.25	1.00	0.28	0.30	0.27
Y5	0.30	0.27	0.24	0.21	0.32	0.34	0.54	0.46	0.28	1.00	0.35	0.59
Y6	0.37	0.35	0.37	0.29	0.36	0.37	0.32	0.29	0.30	0.35	1.00	0.31
Y7	0.21	0.20	0.18	0.16	0.27	0.40	0.58	0.45	0.27	0.59	0.31	1.00

Canonical Correlation Analysis

	Adjusted Canonical Correlation	Approx Canonical Correlation	Squared Standard Error	Canonical Correlation
1	0.553706	0.553073	0.006934	0.306591
2	0.236404	0.234689	0.009442	0.055887
3	0.119186	.	0.009858	0.014205
4	0.072228	.	0.009948	0.005217
5	0.057270	.	0.009968	0.003280

X组的典型变量

	U1	U2	U3	U4	U5
X1	0.4217	0.3429	-0.8577	-0.7884	0.0308
X2	0.19511	-0.6683	0.4434	-0.2691	0.9832
X3	0.1676	-0.8532	-0.2592	0.4688	-0.9141
X4	-0.0229	0.3561	-0.4231	1.0423	0.5244
X5	0.4597	0.7287	0.9799	-0.1682	-0.4392

Y组的典型变量

	V1	V2	V3	V4	V5
Y1	0.4252	-0.0880	0.4918	-0.1284	-0.4823
Y2	0.2089	0.4363	-0.7832	-0.3405	-0.7499
Y3	-0.0359	-0.0929	-0.4778	-0.6059	0.3457
Y4	0.0235	0.9260	-0.0065	0.4044	0.3116
Y5	0.2902	-0.1011	0.2831	-0.4469	0.7030
Y6	0.5157	-0.5543	-0.4125	0.6876	0.1796
Y7	-0.1101	-0.0317	0.9285	0.2739	-0.0141

原始变量与本组典型变量之间的相关系数

	U1	U2	U3	U4	U5
X1	0.8293	0.1093	-0.4853	-0.2469	0.0611
X2	0.7304	-0.4366	0.2001	0.0021	0.4857
X3	0.7533	-0.4661	-0.1056	0.3020	-0.3360
X4	0.6160	0.2225	-0.2053	0.6614	0.3026
X5	0.8606	0.2660	0.3886	0.1484	-0.1246
	V1	V2	V3	V4	V5
Y1	0.7564	0.0446	0.3395	-0.1294	-0.3370
Y2	0.6439	0.3582	-0.1717	-0.3530	-0.3335
Y3	0.3872	0.0373	-0.1767	-0.5348	0.4148
Y4	0.3772	0.7919	-0.0054	0.2886	0.3341
Y5	0.6532	0.1084	0.2092	-0.4376	0.4346
Y6	0.8040	-0.2416	-0.2348	0.4052	0.1964
Y7	0.5024	0.1628	0.4933	-0.1890	0.0678

原始变量与对应组典型变量之间的相关系数

	V1	V2	V3	V4	V5
X1	0.4592	0.0258	-0.0578	-0.0178	0.0035
X2	0.4044	-0.1032	0.0239	0.0002	0.0278
X3	0.4171	-0.1102	-0.0126	0.0218	-0.0192
X4	0.3411	0.0526	-0.0245	0.0478	0.0173
X5	0.4765	0.0629	0.0463	0.0107	-0.0071
	U1	U2	U3	U4	U5
Y1	0.4188	0.0105	0.0405	-0.0093	-0.0193
Y2	0.3565	0.0847	-0.0205	-0.0255	-0.0191
Y3	0.2144	0.0088	-0.0211	-0.0386	0.0238
Y4	0.2088	0.1872	-0.0006	0.0208	0.0191
Y5	0.3617	0.0256	0.0249	-0.0316	0.0249
Y6	0.4452	-0.0571	-0.0280	0.0293	0.0112
Y7	0.2782	0.0385	0.0588	-0.0136	0.0039

可以看出，所有五个表示职业特性的变量与u1有大致相同的相关系数，u1视为形容职业特性的指标。第一对典型变量的第二个变量v1与Y1，Y2，Y5，Y6有较大的相关系数，说明v1主要代表了主管满意度、事业前景满意度、公司地位满意度和工种满意度。而u1和v1之间的相关系数0.5537。

Canonical Redundancy Analysis
Raw Variance of the 'VAR' Variables
Explained by

	Their Own Canonical Variables		The Opposite Canonical Variables	
	Proportion	Cumulative Proportion	Proportion	Cumulative Proportion
1	0.5818	0.5818	0.1784	0.1784
2	0.1080	0.6898	0.0060	0.1844
3	0.0960	0.7858	0.0014	0.1858
<u>4</u>	<u>0.1223</u>	<u>0.9081</u>	0.0006	0.1864
5	0.0919	1.0000	0.0003	0.1867

Raw Variance of the 'WITH' Variables
Explained by

	Their Own Canonical Variables		The Opposite Canonical Variables	
	Proportion	Cumulative Proportion	Proportion	Cumulative Proportion
1	0.3721	0.3721	0.1141	0.1141
2	0.1222	0.4943	0.0068	0.1209
3	0.0740	0.5683	0.0011	0.1220
<u>4</u>	<u>0.1289</u>	<u>0.6972</u>	0.0007	0.1226
5	0.1058	0.8030	0.0003	0.1230

u1和v1解释的本组原始变量的比率：

$$m_{u_1} = \frac{1}{5}(0.8293^2 + 0.7304^2 + \cdots + 0.8606^2) = 0.5818$$

$$n_{v_1} = \frac{1}{7}(0.7564^2 + 0.6439^2 + \cdots + 0.5024^2) = 0.3721$$

X组的原始变量被u1到u5解释了100%

Y组的原始变量被v1到v5解释了80.3%

X组的原始变量被u1到u4解释了90.81%

Y组的原始变量被v1到v4解释了69.72%

六、典型相关系数的显著性检验

- 在进行典型相关分析时，对于两随机向量 (X, Y) ，我们可以提取出 $\min(p, q)$ 对典型变量，问题是进行典型相关分析的目的就是要减少分析变量，简化两组变量间关系分析，提取 $\min(p, q)$ 对变量是否必要？我们如何确定保留多少对典型变量？

若总体典型相关系数 $\lambda_k = 0$ ，则相应的典型变量 U_k ， V_k 之间无相关关系，因此对分析 X 对 Y 的影响不起作用。这样的典型变量可以不予考虑。

巴特莱特（**Bartlett**）提出了一个根据样本数据检验总体典型相关系数 $\lambda_1, \lambda_2, \dots, \lambda_r$ 是否等于零的方法。检验假设为

$$H_0 : \lambda_{k+1} = \lambda_{k+2} = \dots = \lambda_r = 0$$

$$H_1 : \lambda_{k+1} \neq 0$$

用于检验的似然比统计量为：

$$\Lambda_k = \prod_{i=k+1}^r (1 - \hat{\lambda}_i^2)$$

可以证明, $Q_k = -m_k \ln \Lambda_k$ 近似服从 $\chi^2(f_k)$ 分布, 其中自由度

$$f_k = (p - k)(q - k), \quad m_k = (n - k - 1) - \frac{1}{2}(p + q + 1)。$$

我们首先检验 $H_0 : \lambda_1 = \lambda_2 = \cdots = \lambda_r = 0$ 。此时 $k = 0$, 则

$$\Lambda_0 = \prod_{i=1}^r (1 - \hat{\lambda}_i^2) = (1 - \hat{\lambda}_1)(1 - \hat{\lambda}_2) \cdots (1 - \hat{\lambda}_r)$$

$$Q_0 = -m \ln \Lambda_0 = -[(n - 1) - \frac{1}{2}(p + q + 1)] \ln \Lambda_0$$

若 $Q_0 > \chi^2_\alpha(f_0)$ ，则拒绝原假设，也就是说至少有一个典型相关系数大于零，自然应是最大的典型相关系数 $\lambda_1 > 0$ 。

若已判定 $\lambda_1 > 0$ ，则再检验 $H_0 : \lambda_2 = \lambda_3 = \cdots = \lambda_r = 0$ 。此时 $k = 1$ ，则

$$\Lambda_1 = \prod_{i=2}^r (1 - \hat{\lambda}_i^2) = (1 - \hat{\lambda}_2^2)(1 - \hat{\lambda}_3^2) \cdots (1 - \hat{\lambda}_r^2)$$

$$Q_1 = -m_1 \ln \Lambda_1 = -[(n-1-1) - \frac{1}{2}(p+q+1)] \ln \Lambda_1$$

Q_1 近似服从 $\chi^2(f_1)$ 分布，其中 $f_1 = (p-1)(q-1)$ ，如果 $Q_1 > \chi^2_\alpha(f_1)$ ，则拒绝原假设，也即认为 $\lambda_2, \lambda_3, \cdots, \lambda_r$ 至少有一个大于零，自然是 $\lambda_2 > 0$ 。

若已判断 λ_1 和 λ_2 大于零，重复以上步骤直至

$H_0 : \lambda_j = \lambda_{j+1} = \cdots = \lambda_r = 0$ ，此时令

$$\Lambda_{j-1} = \prod_{i=j}^r (1 - \hat{\lambda}_i^2) = (1 - \hat{\lambda}_j)(1 - \hat{\lambda}_{j+1}) \cdots (1 - \hat{\lambda}_r)$$

则

$$Q_{j-1} = -m_{j-1} \ln \Lambda_{j-1} = -[(n-j) - \frac{1}{2}(p+q+1)] \ln \Lambda_{j-1}$$

Q_{j-1} 近似服从 $\chi^2(f_{j-1})$ 分布，

其中 $f_{j-1} = (p-j+1)(q-j+1)$ ，如果 $Q_{j-1} < \chi_{\alpha}^2(f_{j-1})$ ，

则 $\lambda_j = \lambda_{j+1} = \cdots = \lambda_r = 0$ ，于是总体只有 $j-1$ 个典型相关系数不为零，提取 $j-1$ 对典型变量进行分析。

- 例10.1 康复俱乐部对20名中年人测量了三个生理指标：体重(x_1),腰围(x_2), 脉搏(x_3); 三个训练指标：引体向上次数(y_1), 起坐次数(y_2), 跳跃次数(y_3)。分析生理指标与训练指标的相关性。

表 康复俱乐部数据

变量 样本	x_1	x_2	x_3	y_1	y_2	y_3
1	191	36	50	5	162	60
2	189	37	52	2	110	60
3	193	38	58	12	101	101
4	162	35	62	12	105	37
5	189	35	46	13	155	58
6	182	36	56	4	101	42
7	211	38	56	8	101	38

<div> <div>变量</div> <div>样本</div> </div>	x_1	x_2	x_3	y_1	y_2	y_3
8	167	34	60	6	125	40
9	176	31	74	15	200	40
10	154	33	56	17	251	250
11	169	34	50	17	120	38
12	166	33	52	13	210	115
13	154	34	64	14	215	105
14	247	46	50	1	50	50
15	193	36	46	6	70	31
16	202	37	62	12	210	120
17	176	37	54	4	60	25
18	157	32	52	11	230	80
19	156	33	54	15	225	73
20	138	33	68	2	110	43

根据原始数据，计算可得

$$\hat{\Sigma}_{11} = \begin{bmatrix} 579.14 & 65.36 & -61.86 \\ 65.36 & 9.74 & -7.74 \\ -61.86 & -7.74 & 49.39 \end{bmatrix}$$

$$\hat{\Sigma}_{22} = \begin{bmatrix} 26.55 & 218.60 & 127.67 \\ 218.60 & 3718.85 & 2039.64 \\ 127.67 & 2039.64 & 2497.91 \end{bmatrix}$$

$$\hat{\Sigma}_{12} = \begin{bmatrix} -48.32 & -723.63 & -272.18 \\ -8.88 & -122.87 & -29.87 \\ 5.46 & 96.45 & 12.27 \end{bmatrix}$$

$$\hat{\Sigma}_{21} = \begin{bmatrix} -48.32 & -8.88 & 5.46 \\ -723.63 & -122.87 & 96.45 \\ -272.18 & -29.87 & 12.27 \end{bmatrix}$$

$$\hat{\Sigma}_{11}^{-1} = \begin{bmatrix} 0.00723237 & -0.047214 & 0.00165941 \\ -0.047214 & 0.42549329 & 0.00754531 \\ 0.00165941 & 0.00754531 & 0.02350784 \end{bmatrix}$$

$$\hat{\Sigma}_{22}^{-1} = \begin{bmatrix} 0.0732399 & -0.0040789 & -0.00041 \\ -0.0040789 & 0.00071416 & -0.00037 \\ -0.0004126 & -0.0003747 & 0.000727 \end{bmatrix}$$

计算得

$$\hat{\mathbf{A}} = \hat{\Sigma}_{11}^{-1} \hat{\Sigma}_{12} \hat{\Sigma}_{22}^{-1} \hat{\Sigma}_{21} = \begin{bmatrix} -0.2459454 & -0.0551887 & 0.04651367 \\ 4.498811 & 0.90714323 & -0.7392212 \\ -0.0575041 & -0.0138964 & 0.01728371 \end{bmatrix}$$

$$\hat{\mathbf{B}} = \hat{\Sigma}_{22}^{-1} \hat{\Sigma}_{21} \hat{\Sigma}_{11}^{-1} \hat{\Sigma}_{12} = \begin{bmatrix} 0.16178831 & 2.03428439 & 0.223085 \\ 0.04076171 & 0.54877371 & 0.091339 \\ -0.0328274 & -0.4227509 & -0.03208 \end{bmatrix}$$

求得特征值为： $\lambda_1^2 = \mathbf{0.6330}$ ， $\lambda_2^2 = \mathbf{0.0402}$ ， $\lambda_3^2 = \mathbf{0.0053}$ 。

典型相关系数分别为： $\lambda_1 = \mathbf{0.796}$ ， $\lambda_2 = \mathbf{0.201}$ ， $\lambda_3 = \mathbf{0.073}$ 。

$\hat{\mathbf{A}}$ 和 $\hat{\mathbf{B}}$ 相应的特征向量分别为:

$$\alpha_1 = (-0.031, 0.493, -0.008)'$$

$$\alpha_2 = (-0.076, 0.3687, -0.032)'$$

$$\alpha_3 = (-0.008, 0.158, 0.146)'$$

$$\beta_1 = (-0.066, -0.017, 0.014)'$$

$$\beta_2 = (-0.071, 0.002, 0.021)'$$

$$\beta_3 = (-0.245, 0.020, -0.008)'$$

根据前述的典型相关系数显著性检验方法，对于
 $H_0: \lambda_1 = \lambda_2 = \lambda_3 = 0$ ， H_1 : 至少有一个不为零。

$$\Lambda_0 = \prod_{i=1}^3 (1 - \hat{\lambda}_i^2)$$

$$= (1 - 0.632992)(1 - 0.040223)(1 - 0.005266)$$

$$= 0.35039053$$

$$Q_0 = -m \ln \Lambda_0 = -[(n-1) - \frac{1}{2}(p+q+1)] \ln \Lambda_0$$

$$= -[(20-1) - \frac{1}{2}(3+3+1)] \ln \Lambda_0$$

$$= -15.5 \ln \Lambda_0 = 16.255$$

$Q_0 < \chi_{0.05}^2(9) = 16.919$, 故在 $\alpha=0.05$ 下, 生理指标与训练指标之间不存在相关性;

而在 $\alpha=0.10$ 下, $Q_0 > \chi_{0.10}^2(9) = 14.684$, 生理指标与训练指标之间存在相关性, 且第一对典型变量相关性显著。
继续检验:

$$\Lambda_1 = \prod_{i=2}^3 (1 - \hat{\lambda}_i^2) = (1 - 0.040223)(1 - 0.005266) = 0.95472266$$

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

$$= -[(20-2) - \frac{1}{2}(3+3+1)] \ln \Lambda_1$$

$$= -14.5 \ln \Lambda_1 = 0.672$$

$Q_1 < \chi_{0.10}^2(4) = 7.779$ ，故在 $\alpha = 0.10$ 下，第二对典型变量间相关性不显著。说明生理指标和训练指标之间只有一对典型变量，即：

$$U_1 = -0.031x_1 + 0.493x_2 - 0.008x_3$$

$$V_1 = -0.066y_1 - 0.017y_2 + 0.014y_3$$

从相关矩阵出发计算典型相关

- 为消除量纲影响，对数据先做标准化变换，然后再做典型相关分析。
- 显然，经标准化变换之后的协差阵就是相关系数矩阵，因而，也即通常应从相关矩阵出发进行典型相关分析。

- 对于例10.1从相关系数矩阵出发进行典型相关分析。

$$\hat{\mathbf{R}}_{11} = \begin{bmatrix} 1 & & \\ 0.87024349 & 1 & \\ -0.365762 & -0.3528921 & 1 \end{bmatrix}$$

$$\hat{\mathbf{R}}_{22} = \begin{bmatrix} 1 & & \\ 0.69572742 & 1 & \\ 0.49576018 & 0.66920608 & 1 \end{bmatrix}$$

$$\hat{\mathbf{R}}_{12} = \begin{bmatrix} -0.3896937 & -0.4930836 & -0.2262956 \\ -0.5522321 & -0.645598 & -0.1914994 \\ 0.15064802 & 0.22503808 & 0.03493306 \end{bmatrix} = \hat{\mathbf{R}}'_{21}$$

$$\hat{\mathbf{A}}_z = \hat{\mathbf{R}}_{11}^{-1} \hat{\mathbf{R}}_{12} \hat{\mathbf{R}}_{22}^{-1} \hat{\mathbf{R}}_{21} = \begin{bmatrix} -0.2459455 & -0.4255619 & 0.15927699 \\ 0.58342554 & 0.9071432 & -0.3282724 \\ -0.0167929 & -0.0312927 & 0.01728372 \end{bmatrix}$$

$$\hat{\mathbf{B}}_z = \hat{\mathbf{R}}_{22}^{-1} \hat{\mathbf{R}}_{21} \hat{\mathbf{R}}_{11}^{-1} \hat{\mathbf{R}}_{12} = \begin{bmatrix} 0.16178827 & 0.1718776 & 0.022998202 \\ 0.48244159 & 0.5487737 & 0.111448282 \\ -0.3184294 & -0.3464725 & -0.032080511 \end{bmatrix}$$

计算得 $\hat{\mathbf{A}}_z, \hat{\mathbf{B}}_z$ 的特征值为: $\lambda_1^2 = 0.6330$, $\lambda_2^2 = 0.0402$,

$\lambda_3^2 = 0.0053$ 。其结果同从协差阵出发计算的特征值相同,

因此检验结果也相同, 提取第一典型变量, 按照类似的方法可求得典型变量系数向量:

$$\alpha_1^* = (-0.7754, 1.5793, -0.0591)'$$

$$\alpha_2^* = (-1.8844, 1.1806, -0.2311)'$$

$$\alpha_3^* = (-0.1910, 0.5060, 1.0508)'$$

$$\beta_1^* = (-0.3495, -1.0540, 0.7164)'$$

$$\beta_2^* = (-0.3755, 0.1235, 1.0622)'$$

$$\beta_3^* = (-1.2966, 1.2368, -0.4188)'$$

可得到标准化的第一对典型变量：

$$U_1^* = -0.7754x_1^* + 1.5793x_2^* - 0.0591x_3^*$$

$$V_1^* = 0.3495y_1^* - 1.0540y_2^* + 0.7164y_3^*$$

典型载荷分析

VAR 变量 及其典型变量之间的相关性				
		U1	U2	U3
x1	体重	0. 6206	-0. 7724	-0. 1350
x2	腰围	0. 9254	-0. 3777	-0. 0310
x3	脉搏	-0. 3328	0. 0415	0. 9421

WITH 变量 及其典型变量之间的相关性				
		V1	V2	V3
y1	引体向上	-0. 7276	0. 2370	-0. 6438
y2	起坐次数	-0. 8177	0. 5730	0. 0544
y3	跳跃次数	-0. 1622	0. 9586	-0. 2339

典型冗余分析

通过以下变量解释的 VAR 变量 原始方差

典型变量号	它们自己的典型变量		典型R 方	对立面典型变量	
	比例	累积比例		比例	累积比例
1	0.3712	0. 3712	0. 6330	0. 2349	0. 2349
2	0.5436	0. 9148	0. 0402	0. 0219	0. 2568
3	0.0852	1. 0000	0. 0053	0. 0004	0. 2573

通过以下变量解释的 WITH 变量 原始方差

典型变量号	它们自己的典型变量		典型R 方	对立面典型变量	
	比例	累积比例		比例	累积比例
1	0.4111	0. 4111	0. 6330	0. 2602	0. 2602
2	0.5635	0. 9746	0. 0402	0. 0227	0. 2829
3	0.0254	1. 0000	0. 0053	0. 0001	0. 2830

典型冗余分析

通过以下变量解释的 VAR 变量标准化方差

典型变量号	它们自己的典型变量		典型R 方	对立面典型变量	
	比例	累积比例		比例	累积比例
1	0. 4508	0. 4508	0. 6330	0. 2854	0. 2854
2	0. 2470	0. 6978	0. 0402	0. 0099	0. 2953
3	0. 3022	1. 0000	0. 0053	0. 0016	0. 2969

通过以下变量解释的 WITH 变量标准化方差

典型变量号	它们自己的典型变量		典型R 方	对立面典型变量	
	比例	累积比例		比例	累积比例
1	0. 4081	0. 4081	0. 6330	0. 2584	0. 2584
2	0. 4345	0. 8426	0. 0402	0. 0175	0. 2758
3	0. 1574	1. 0000	0. 0053	0. 0008	0. 2767

载荷 ²	U1	U2	U3	方差
x1	0.3851	0.5966	0.0182	609.6210
x2	0.8564	0.1427	0.0010	10.2526
x3	0.1108	0.0017	0.8876	51.9895
总和				671.8631

$$\frac{0.3851 \times 609.6201 + 0.8504 \times 10.2526 + 0.1108 \times 51.9895}{609.6201 + 10.2526 + 51.9895} = 0.3712$$

$$\frac{0.5966 \times 609.6201 + 0.1427 \times 10.2526 + 0.0017 \times 51.9895}{609.6201 + 10.2526 + 51.9895} = 0.5436$$

$$\frac{0.0182 \times 609.6201 + 0.0010 \times 10.2526 + 0.8876 \times 51.9895}{609.6201 + 10.2526 + 51.9895} = 0.0852$$

载荷 ²	V1	V2	V3	方差
y1	0.5294	0.0562	0.4145	27.9447
y2	0.6686	0.3283	0.0030	3914.5763
y3	0.0263	0.9189	0.0547	2629.3789
总和				6571.9000

$$\frac{0.5294 \times 27.9447 + 0.6686 \times 3914.5763 + 0.0263 \times 2629.3789}{6571.90} = 0.4111$$

$$\frac{0.0562 \times 27.9447 + 0.3283 \times 3914.5763 + 0.9189 \times 2629.3789}{6571.90} = 0.5635$$

$$\frac{0.4145 \times 27.9447 + 0.003 \times 3914.5763 + 0.0547 \times 2629.3789}{6571.90} = 0.0254$$

VAR 变量 和 WITH 变量 的典型变量之间的相关性					载荷 ²		
		V1	V2	V3	1	2	3
x1	体重	0.4938	-0.1549	-0.0098	0.2438	0.0240	0.0001
x2	腰围	0.7363	-0.0757	-0.0022	0.5421	0.0057	0.0000
x3	脉搏	-0.2648	0.0083	0.0684	0.0701	0.0001	0.0047
WITH 变量 和 VAR 变量 的典型变量之间的相关性					载荷 ²		
		U1	U2	U3	1	2	3
y1	引体向上	-0.5789	0.0475	-0.0467	0.3351	0.0023	0.0022
y2	起坐次数	-0.6506	0.1149	0.004	0.4233	0.0132	0.0000
y3	跳跃次数	-0.129	0.1923	-0.017	0.0166	0.0370	0.0003

典型冗余分析

VAR 变量 和 WITH 变量 前 M 个典型变量之间的多重相关系数平方				
M		1	2	3
x1	体重	0. 2438	0. 2678	0. 2679
x2	腰围	0. 5421	0. 5478	0. 5478
x3	脉搏	0. 0701	0. 0702	0. 0749

WITH 变量 和 VAR 变量 前 M 个典型变量之间的多重相关系数平方				
M		1	2	3
y1	引体向上	0. 3351	0. 3374	0. 3396
y2	起坐次数	0. 4233	0. 4365	0. 4365
y3	跳跃次数	0. 0167	0. 0536	0. 0539

SAS 程序

```
Proc cancorr data=SAS数据集 out= SAS数据集  
outstat= SAS数据集 all;
```

with 变量名;

列出被分析的两组变量中的**第二组**变量

Var 变量名;

列出被分析的两组变量中的**第一组**变量

Run;

produce simple statistics, input variable
correlations, and canonical redundancy
analysis

OUT=

包含原始数据和典型变量得分的数据集

OUTSTAT=

包含分析中用到的各种统计量


SAS 程序

```
Proc cancorr data=SAS数据集 out= SAS数据集  
outstat= SAS数据集 all vdep / vreg;
```

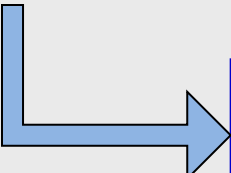
with 变量名;

Var 变量名;

Run;



以Var变量为因变量，
以with变量为自变
量，进行多元回归
分析



以with变量为因变量，
以Var变量为自变量，
进行多元回归分析



浙江工商大学
ZHEJIANG GONGSHANG UNIVERSITY

Thank You