重慶工商大學

CHONGQING TECHNOLOGY AND BUSINESS UNIVERSITY

典型相关分析

CANONICAL CORRELATION ANALYSIS

数学与统计学院 杨炜明

本 章 要 点

皮尔逊相关系数—两个变量的相关。

什么是典型相关分析？典型相关—两组变量的相关。

如何求典型相关系数。

典型相关系数是什么变量的相关系数？原始变量还是新变量？

如果是新变量，那么新变量如何求得。

典型相关分析与主成分分析有何相似之处。

简单相关，复相关，典型相关之间的关系。

第一节 引言

有许多分析变量间关系的方法，它们各有千秋，但是我们还没有学习过关于两组变量间关系的分析，这一章我们讨论关于变量间的相关关系。为了这一章的学习，也为了我们的学习更多的相关分析的方法，下面我们现对变量的相关分析进行总结。

相关分析是分析变量之间的相关关系。

（一）简单相关系数

$$\gamma = \frac{\sum (Y - Y)(X_1 - \bar{X}_1)}{\sqrt{\sum (\mathrm{Y}\text{-}\bar{\mathrm{Y}})^2} \sqrt{\sum (X_1 - \bar{X})^2}}$$

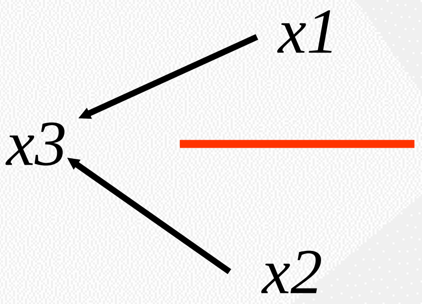
（二）偏相关系数

1、问题提出

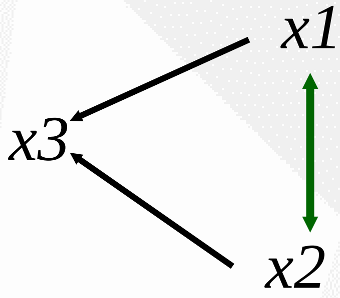
简单相关系数在一定情况下无法较为真实准确地度量事务之间的相关关系，往往有夸大的倾向。例如：在研究商品的需求量、价格和消费者收入时会发现，需求量和价格之间的关系实际上还包含了消费者收入对商品需求量的影响。

重慶工商大學
CHONGQING TECHNOLOGY AND BUSINESS UNIVERSITY

设被解释变量x3受两个彼此独立的自变量x1和x2的影响。



若自变量x1和x2彼此不独立，存在一定的相关关系：



这时，产生了通径

$$\begin{cases} x_1 \rightarrow x_2 \rightarrow y \\ x_2 \rightarrow x_1 \rightarrow y \end{cases}$$

2、偏相关系数的定义

设x1，x2，y是三个变量，如果要计算x2给定的条件下， x1 和y的相关系数，应该用偏相关系数更合理，那么偏相关系数为：

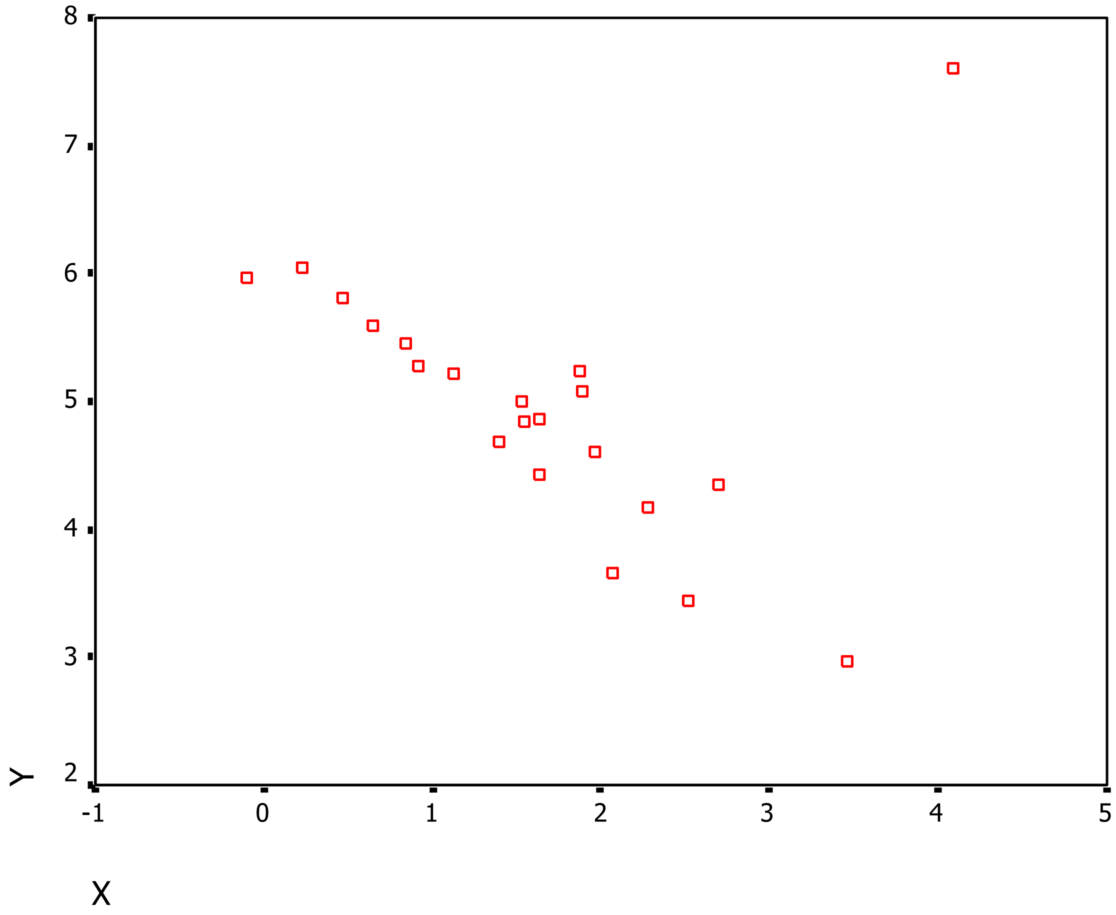
$$r_{y,x_1\cdot x_2}=\frac{r_{(y,x_1)}-r_{(y,x_2)}r_{(x_1,x_2)}}{\sqrt{1-r^2_{(y,x_2)}}\sqrt{1-r^2_{(x_1,x_2)}}}$$

重庆工商大学 CHONGQING TECHNOLOGY AND BUSINESS UNIVERSITY

为什么说皮尔逊相关系数的不稳健的呢？下面讨论一个例子，将进一步说明皮尔逊相关系数的缺点。下表的数据是模拟数据：

x	y	x	y
.84	5.46	3.46	2.97
.46	5.80	.22	6.05
1.64	4.87	1.39	4.69
2.07	3.66	1.13	5.22
2.27	4.18	1.88	5.23
-.10	5.97	1.97	4.60
4.09	7.60	1.53	5.00
.65	5.60	.92	5.27
2.52	3.45	1.89	5.07
2.70	4.34	1.64	4.42
1.54	4.84		

其散点图如下：



Pearson相关系数表

X	Pearson Correlation	1	-.283
	Sig. (2-tailed)	.	.214
Y	Pearson Correlation	-.283	1
	Sig. (2-tailed)	.214	.

该表的含义是说**Pearson相关系数为-0.283**，Sig.是检验Ho: X和Y不相关的显著性水平（P值），因为Sig.＝0.214，则不能拒绝原假设，但是从前面的散点图，实际上除了那个异常点外，二者是相关的。

（三）Spearman相关系数

在给定一列数对（x_1,y_1），…，（x_n,y_n）之后，要检验他们所代表的二元变量X和Y是否相关。首先将X和Y的观测值分别排序，分别得各自得秩统计量，Spearman相关检验的含义是直接对秩统计量计算相关系数，即计算R和S的相关系数：

$$(R_1,S_1),\cdots,(R_n,S_n)$$

$$r_s = \frac{\sum_{i=1}^{n}(R_i - \bar{R})(S_i - \bar{S})}{\sqrt{\sum_{i=1}^{n}(R_i - \bar{R})^2 \sum_{i=1}^{n}(S_i - \bar{S})^2}} = 1 - \frac{6\sum_{i=1}^{n} d_i^2}{n(n^2 - 1)}$$

$$D_i = R_i - S_i$$

优点：稳健，不受极端值影响

（四）复相关系数

简单相关系数和偏相关系数实际上均是讨论两个变量的关系，但常常我们会讨论一个变量和一组变量的相关，这叫复相关系数。实际上一个变量和一组变量的复相关是以这个变量为被解释变量，以这组变量为回归因子，建立回归模型的可决系数R^2.

重庆工商大学
CHONGQING TECHNOLOGY AND BUSINESS UNIVERSITY

如何讨论两组变量的关系呢？

通常情况下，为了研究两组变量

$$(x_1, x_2, \cdots, x_p) \qquad (y_1, y_2, \cdots, y_q)$$

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，这样又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。

在解决实际问题中，这种方法有广泛的应用。如，在工厂里常常要研究产品的p个质量指标 $(x_1, x_2, \cdots, x_p)$ 和q个原材料的指标$(y_1, y_2, \cdots, y_q)$ 之间的相关关系；也可以是采用典型相关分析来解决的问题。如果能够采用类似于主成分的思想，分别找出两组变量的线性组合既可以使变量个数简化，又可以达到分析相关性的目的。

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个职业满意变量。讨论 两组指标之间是否相联系。

X组:	Y组:
X1—用户反馈	Y1—主管满意度
X2—任务重要性	Y2—事业前景满意度
X3—任务多样性	Y3—财政满意度
X4—任务特殊性	Y4—工作强度满意度
X5—自主权	Y5—公司地位满意度
	Y6—工作满意度
	Y7—总体满意度

重庆工商大学
CHONGQING TECHNOLOGY AND BUSINESS UNIVERSITY

	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

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（两两相关）

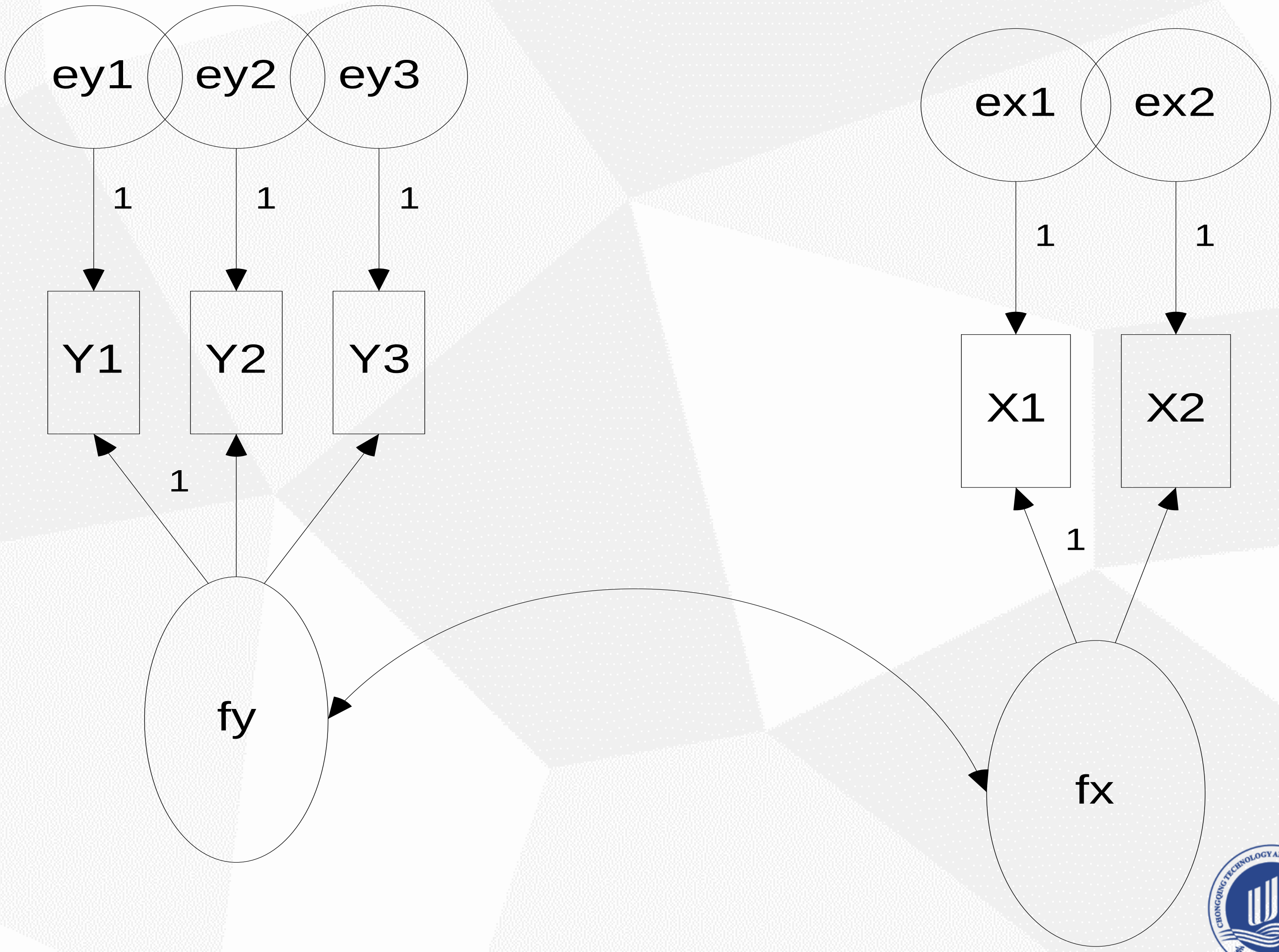
	x1	x2	y1	y2	y3
x1	1.00	0.80	0.26	0.67	0.34
x2	0.80	1.00	0.33	0.59	0.34
y1	0.26	0.33	1.00	0.37	0.21
y2	0.67	0.59	0.37	1.00	0.35
y3	0.34	0.34	0.21	0.35	1.00

典型相关分析方法(canonical correlation analysis---CCA)最早源于荷泰林(H，Hotelling)于1936年在《生物统计》期刊上发表的一篇论文《两组变式之间的关系》。他所提出的方法经过多年的应用及发展，逐渐达到完善，在70年代臻于成熟。

由于典型相关分析涉及较大量的矩阵计算，其方法的应用在早期曾受到相当的限制。但随着当代计算机技术及其软件的迅速发展，弥补了应用典型相关分析中的困难，因此它的应用开始走向普及化。

重慶工商大學
CHONGQING TECHNOLOGY AND BUSINESS UNIVERSITY

典型相关分析是结构方程模型的特例



$$\begin{cases} u_1 = a_{11}x_1 + a_{21}x_2 \\ V_1 = b_{11}y_1 + b_{21}y_2 + b_{31}y_3 \end{cases}$$

$$\rho(u_1, v_1) = ?$$

x_1

x_2

$$\begin{cases} u_2 = a_{12}x_1 + a_{22}x_2 \\ v_2 = b_{12}y_1 + b_{22}y_2 + b_{32}y_3 \end{cases}$$

$$\rho(u_2, v_2) = ?$$

y_1

y_2

y_3

第二节　总体典型相关分析

一、典型相关和典型相关变量的定义

设X和Y分别为p和q维的随机向量。如果存在a_1和b_1，使得

$$\rho(\mathbf{a}_1'\mathbf{X},\mathbf{b}_1'\mathbf{Y})=\max_{Var(\boldsymbol{\alpha}_1'\mathbf{X})=1,Var(\boldsymbol{\beta}_1'\mathbf{Y})=1}\rho(\boldsymbol{\alpha}_1'\mathbf{X},\boldsymbol{\beta}_1'\mathbf{Y})$$

则称 $u_1=\mathbf{a}_1'\mathbf{X},v_1=\mathbf{b}_1'\mathbf{Y}$ 是X和Y的第一对典型相关变量，其相关系数称为典型相关系数。不妨假设$\rho>0$。

如果存在，a_k和b_k（k=2，3，…，m）。满足以下三个条件。

(1）第k对典型变量与前面k-1对典型变量都不相关；

(2）其典型变量的方差均为1；

(3）第1对典型变量的相关系数的绝对值最大。后面依次减少。

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，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。如此下去，直至两组变量的相关性被提取完为止。

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

$$\cdots$$

$$\begin{cases} u_r = a_{1r}x_1 + a_{2r}x_2 + \cdots + a_{pr}x_p \\ v_r = b_{1r}y_1 + b_{2r}y_2 + \cdots + b_{qr}y_q \end{cases}$$

u_2和v_1，u_1和v_2， u_2和u_1， v_1和v_2相互无关，但u_2和v_2相关。如此继续下去，直至进行到r步，$r \le \min(p,q)$，可以得到r组变量。

$$U = (u_1, \cdots, u_r)'$$

$$V = (v_1, \cdots, v_r)'$$

从而达到降维的目的。

二、典型相关的解法

考虑两组变量的向量

$$\mathbf{Z} = (x_1, x_2, \cdots, x_p, y_1, y_2, \cdots, y_q)'$$

其协方差阵为

$$\mathbf{\Sigma} = \begin{array}{cc} \begin{bmatrix} \mathbf{\Sigma}_{11} & \mathbf{\Sigma}_{12} \\ \mathbf{\Sigma}_{21} & \mathbf{\Sigma}_{22} \end{bmatrix} & \begin{array}{c} p \\ q \end{array} \\ \begin{array}{cc} \quad p & \qquad q \end{array} & \end{array}$$

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；Σ_{12}和Σ_{21}是X和Y的其协方差矩阵。

1. 第一对典型相关变量的解法

$$u_1 = \boldsymbol{\alpha}_1'\mathbf{X} \qquad v_1 = \boldsymbol{\beta}_1'\mathbf{Y}$$

求第一对典型变量相关变量就等价于，求

$$\boldsymbol{\alpha}_1 = (\alpha_{11}, \alpha_{21}, \cdots, \alpha_{p1})' \in \mathbf{R}^p$$

$$\boldsymbol{\beta}_1 = (\beta_{11}, \beta_{21}, \cdots, \beta_{q1})' \in \mathbf{R}^q$$

满足条件以下条件下

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

$$Var(v_1) = \boldsymbol{\beta}_1' Var(\mathbf{Y}) \boldsymbol{\beta}_1 = \boldsymbol{\beta}_1' \boldsymbol{\Sigma}_{22} \boldsymbol{\beta}_1 = 1$$

重慶工商大學
CHONGQING TECHNOLOGY AND BUSINESS UNIVERSITY

$$\rho_{u_1,v_1} = Cov(u_1, v_1)$$

$$= \boldsymbol{\alpha}_1' Cov(X, Y)\boldsymbol{\beta}_1 = \boldsymbol{\alpha}_1'\boldsymbol{\Sigma}_{12}\boldsymbol{\beta}_1$$

$$= \max$$

可见典型相关分析就是求a_1和b_1，即线性组合的系数，使二者的相关系数ρ_1达到最大，假设$\rho_1>0$。

重庆工商大学 CHONGQING TECHNOLOGY AND BUSINESS UNIVERSITY

这是一个求条件极值的问题，我们用拉格朗日乘数法，有其目标函数为

$$\varphi(\boldsymbol{\alpha}_1,\boldsymbol{\beta}_1)=\boldsymbol{\alpha}_1' Var(\mathbf{Y})\boldsymbol{\beta}_1-\frac{\lambda_1}{2}\boldsymbol{\alpha}_1' Var(\mathbf{Y})\boldsymbol{\alpha}_1-\frac{\lambda_2}{2}\boldsymbol{\beta}_1' Var(\mathbf{Y})\boldsymbol{\beta}_1$$

分别对目标函数关于α_1和 β_1求导，并令其倒数为零，有

$$\begin{cases}\partial\dfrac{\varphi(\boldsymbol{\alpha}_1,\boldsymbol{\beta}_1)}{\partial\boldsymbol{\alpha}_1}=\boldsymbol{\Sigma}_{12}\boldsymbol{\beta}_1-\lambda_1\boldsymbol{\Sigma}_{11}\boldsymbol{\alpha}_1=0\\ \partial\dfrac{\varphi(\boldsymbol{\alpha}_1,\boldsymbol{\beta}_1)}{\partial\boldsymbol{\beta}_1}=\boldsymbol{\Sigma}_{21}\boldsymbol{\alpha}_1-\lambda_1\boldsymbol{\Sigma}_{22}\boldsymbol{\beta}_1=0\end{cases}$$

重慶工商大學
CHONGQING TECHNOLOGY AND BUSINESS UNIVERSITY

分别用α_1和 β_1左乘方程第一式和第二式，则有

$$\lambda_1 = \lambda_2 = \boldsymbol{\alpha}_1' \boldsymbol{\Sigma}_{12} \boldsymbol{\beta} = \rho(u,v) = \lambda$$

进而，我们用$\boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}$左乘正规方程的第二式，有

$$\boldsymbol{\Sigma}_{12}\boldsymbol{\beta}_1=\frac{1}{\lambda}\boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21}\boldsymbol{\alpha}_1$$

并将其代入第一式，得

$$\boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21}\boldsymbol{\alpha}_1-\lambda^2\boldsymbol{\Sigma}_{11}\boldsymbol{\alpha}_1=0$$

$$\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21}\boldsymbol{\alpha}_1-\lambda^2\boldsymbol{\alpha}_1=0$$

$$\left(\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21}-\lambda^2\mathbf{I}_p\right)\boldsymbol{\alpha}_1=0$$

重慶工商大學
CHONGQING TECHNOLOGY AND BUSINESS UNIVERSITY

同理可得，

$$\left(\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{12}-\lambda^{2}\mathbf{I}_{q}\right)\boldsymbol{\beta}_{1}=0$$

由此，为了求典型相关系数与$\boldsymbol{\alpha}_1$和 $\boldsymbol{\beta}_1$，就是求解方程

$$\begin{cases}\left|\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21}-\lambda^{2}\mathbf{I}_{p}\right|=0\\ \left|\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{12}-\lambda^{2}\mathbf{I}_{q}\right|=0\end{cases}$$

求其特征根λ^2，再求其对应的特征向量α_1和β_1。

令 $$\Sigma_{11}^{-1}\Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}=M_1$$

$$\Sigma_{22}^{-1}\Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12}=M_2$$

可以证明，M1和M2有相同的非零特征根。

$$M_1=\Sigma_{11}^{-1/2}\Sigma_{11}^{-1/2}\Sigma_{12}\Sigma_{22}^{-1/2}\Sigma_{22}^{-1/2}\Sigma_{21}=\mathbf{AB}$$

$$\mathbf{BA}=\Sigma_{11}^{-1/2}\Sigma_{12}\Sigma_{22}^{-1/2}\Sigma_{22}^{-1/2}\Sigma_{21}\Sigma_{11}^{-1/2}$$

令 $$\mathbf{T}=\Sigma_{11}^{-1/2}\Sigma_{12}\Sigma_{22}^{-1/2}$$

$$\mathbf{BA}=\Sigma_{11}^{-1/2}\Sigma_{12}\Sigma_{22}^{-1/2}\Sigma_{22}^{-1/2}\Sigma_{21}\Sigma_{11}^{-1/2}=\mathbf{TT}'$$

M1与TT'有相同的非零特征根。同理有

$$M_2 = \Sigma_{22}^{-1/2}\Sigma_{22}^{-1/2}\Sigma_{21}\Sigma_{11}^{-1/2}\Sigma_{11}^{-1/2}\Sigma_{12} = \mathbf{CD}$$

$$\mathbf{DC} = \Sigma_{22}^{-1/2}\Sigma_{21}\Sigma_{11}^{-1/2}\Sigma_{11}^{-1/2}\Sigma_{12}\Sigma_{22}^{-1/2}$$

令 $$\mathbf{T} = \Sigma_{11}^{-1/2}\Sigma_{12}\Sigma_{22}^{-1/2}$$

$$\mathbf{DC} = \Sigma_{22}^{-1/2}\Sigma_{21}\Sigma_{11}^{-1/2}\Sigma_{11}^{-1/2}\Sigma_{12}\Sigma_{22}^{-1/2} = \mathbf{T'T}$$

M2与T' T有相同的非零特征根。

重庆工商大学
CHONGQING TECHNOLOGY AND BUSINESS UNIVERSITY

所以TT'和T' T 有相同的非零特征根。由上面的分析，二者的相同非零特征根个数最多为p个（因p≤q）。M1的p个非零特征根依次为　　　　　　　　　　，对应的特征向量为

$$\lambda_1^2 \geq \ldots \geq \lambda_p^2 \geq 0$$

$$\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_p$$

M2的p个非零特征根依次对应的特征向量为

$$\mathbf{b}_1, \mathbf{b}_2, \ldots, \mathbf{b}_p$$

重慶工商大學
CHONGQING TECHNOLOGY AND BUSINESS UNIVERSITY

2. 典型相关系数和典型变量的一般求法

求TT’的非零特征根。$\lambda_1^2 \geq \lambda_2^2 \geq \cdots \geq \lambda_p^2 > 0$

设其正交特征向量分别为$\mathbf{t}_1, \ldots, \mathbf{t}_m$

$$\mathbf{a}_k = \boldsymbol{\Sigma}_{11}^{-1/2} \mathbf{t}_k \qquad \mathbf{b}_k = \frac{1}{\lambda_1} \boldsymbol{\Sigma}_{22}^{-1/2} \boldsymbol{\Sigma}_{21} \mathbf{a}_k$$

则 $u_k = \mathbf{a}_\mathbf{k}' \mathbf{X} \quad V_k = \mathbf{b}_\mathbf{k}' \mathbf{Y}$

为第k对典型变量，其相关系数为λ_k。

最大的特征根对应的特征向量构成的第一对线性组合，有最大的相关系数λ_1。

$$
\begin{aligned}
\operatorname{cov}(\mathbf{a}_1'\mathbf{x},\mathbf{b}_1'\mathbf{y}) &= \mathbf{a}_1'\boldsymbol{\Sigma}_{12}\mathbf{b}_1 \\
&= \left(\boldsymbol{\Sigma}_{11}^{-1/2}\mathbf{t}_1\right)'\boldsymbol{\Sigma}_{12}\left(\frac{1}{\lambda_1}\boldsymbol{\Sigma}_{22}^{-1/2}\boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1/2}\mathbf{t}_1\right) \\
&= \frac{1}{\lambda_1}\mathbf{t}_1{}'\boldsymbol{\Sigma}_{11}^{-1/2}\boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1/2}\boldsymbol{\Sigma}_{22}^{-1/2}\boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1/2}\mathbf{t}_1 \\
&= \frac{1}{\lambda_1}\mathbf{t}_1'\mathbf{T}\mathbf{T}'\mathbf{t}_1 \\
&= \frac{1}{\lambda_1}\mathbf{t}_1'\lambda_1^2\mathbf{t}_1 = \lambda_1\mathbf{t}_1'\mathbf{t}_1 = \lambda_1
\end{aligned}
$$

这是因为：

$$\left[\operatorname{cov}(\mathbf{a}_1'\mathbf{x},\mathbf{b}_1'\mathbf{y})\right]=\left[\mathbf{a}_1'\boldsymbol{\Sigma}_{12}\mathbf{b}_1\right]$$

$$=\left(\boldsymbol{\Sigma}_{11}^{-1/2}\mathbf{t}_1\right)'\boldsymbol{\Sigma}_{12}\left(\frac{1}{\lambda_1}\boldsymbol{\Sigma}_{22}^{-1/2}\boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1/2}\mathbf{t}_1\right)$$

$$=\frac{1}{\lambda_1}\mathbf{t}_1'\boldsymbol{\Sigma}_{11}^{-1/2}\boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1/2}\boldsymbol{\Sigma}_{22}^{-1/2}\boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1/2}\mathbf{t}_1$$

$$=\frac{1}{\lambda_1}\mathbf{t}_1'\mathbf{T}\mathbf{T}'\mathbf{t}_1$$

$$=\frac{1}{\lambda_1}\left(\sum_{i=1}^{p}\lambda_i^2\mathbf{t}_1'\mathbf{t}_i\mathbf{t}_i'\mathbf{t}_1\right)\le\frac{1}{\lambda_1}\left(\sum_{i=1}^{p}\lambda_1^2\mathbf{t}_1'\mathbf{t}_i\mathbf{t}_i'\mathbf{t}_1\right)=\lambda_1$$

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。

重庆工商大学
CHONGQING TECHNOLOGY AND BUSINESS UNIVERSITY

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：

$$u_2 = \mathbf{a}_2'\mathbf{x} \qquad v_2 = \mathbf{b}_2'\mathbf{y}$$

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

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

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

$$\operatorname{cov}(v_1, v_2) = \operatorname{cov}(\mathbf{b}_1'\mathbf{y}, \mathbf{b}_2'\mathbf{y}) = \mathbf{b}_1'\mathbf{\Sigma}_{11}\mathbf{b}_2 = 0$$

重庆工商大学 CHONGQING TECHNOLOGY AND BUSINESS UNIVERSITY

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

实际上，两组变量的典型变量共有r=min(p,q)对。有

$$\mathbf{a}_k = \mathbf{\Sigma}_{11}^{-1/2}\mathbf{t}_k \qquad \mathbf{b}_k = \lambda_k^{-1}\mathbf{\Sigma}_{22}^{-1/2}\mathbf{\Sigma}_{21}\mathbf{a}_k$$

$$k = 1, 2, 3, \cdots, r$$

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：

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

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	调整典型相关系数	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	0.7689	-1.4787
X2	0.2721	1.6443
Y组典型变量的系数		
	V1	V2
Y1	0.0491	1.0003
Y2	0.8975	-0.5837
Y3	0.1900	0.2956

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

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

```
data=read.table("c:/can.txt")
names(data)=c("x1","x2","x3","x4","x5","x6","y1","y2","y3","y4","y5")
can=cancor(data[,1:6],data[7:11])
```

28年的宏观经济数据，其中X1-X6为经济数据，Y1-Y5为投资数据。

重慶工商大學 CHONGQING TECHNOLOGY AND BUSINESS UNIVERSITY

三、典型变量的性质

1、相关系数

此处的相关系数包括：

不同组同对典型变量之间的相关系数（相关）；

不同组不同对典型变量之间的相关系数（不相关）；

同组的典型变量之间的关系（不相关）。

$$u_k = \mathbf{a}'_k \mathbf{x} \qquad v_k = \mathbf{b}'_k \mathbf{y} \qquad k, i = 1, 2, \cdots, r; k \neq i$$

1. X组的典型变量之间互不相关：

$$\begin{aligned}
&\operatorname{cov}(u_k, u_i) \\
&= \operatorname{cov}(\mathbf{a}'_\mathbf{k}\mathbf{x}, \mathbf{a}'_\mathbf{i}\mathbf{x}) \\
&= \mathbf{a}'_k \boldsymbol{\Sigma}_{11} \mathbf{a}_i \\
&= \mathbf{a}'_k \boldsymbol{\Sigma}_{11}^{1/2} \boldsymbol{\Sigma}_{11}^{1/2} \mathbf{a}_i \\
&= \mathbf{t}'_k \mathbf{t}_i = 0
\end{aligned}$$

由于其特征向量是正交的。

2. Y组的典型变量之间是互不相关

$$
\begin{aligned}
&\operatorname{cov}(v_k, v_i) \\
&= \operatorname{cov}(b_{\mathbf{k}}'\mathbf{y}, b_{\mathbf{i}}'\mathbf{y}) \\
&= \lambda_k^{-1}\mathbf{a}_k'\boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{22}\lambda_i^{-1}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21}\mathbf{a}_i \\
&= \frac{1}{\lambda_k\lambda_i}\mathbf{t}_k'\mathbf{T}\mathbf{T}'\mathbf{t}_i \\
&= \frac{1}{\lambda_k\lambda_i}\lambda_i\mathbf{t}_k'\mathbf{t}_i \\
&= 0
\end{aligned}
$$

重庆工商大学 CHONGQING TECHNOLOGY AND BUSINESS UNIVERSITY

3. 典型变量之间的相关性

不同组内典型变量之间的相关系数为

$$\mathrm{cov}(u_i, v_j) = \mathrm{cov}(\mathbf{a}_i' x, \mathbf{b}_i' y)$$

$$= \mathbf{t}_\mathbf{i}' \boldsymbol{\Sigma}_{11}^{-1/2} \boldsymbol{\Sigma}_{\mathbf{12}} \lambda_j^{-1} \boldsymbol{\Sigma}_{22}^{-1} \boldsymbol{\Sigma}_{21} \boldsymbol{\Sigma}_{11}^{-1/2} \mathbf{t}_j$$

$$= \lambda_j^{-1} \mathbf{t}_\mathbf{i}' \boldsymbol{\Sigma}_{11}^{-1/2} \boldsymbol{\Sigma}_{\mathbf{12}} \boldsymbol{\Sigma}_{22}^{-1} \boldsymbol{\Sigma}_{21} \boldsymbol{\Sigma}_{11}^{-1/2} \mathbf{t}_j$$

$$= \lambda_j^{-1} \mathbf{t}_\mathbf{i}' \boldsymbol{\Sigma}_{11}^{-1/2} \boldsymbol{\Sigma}_{\mathbf{12}} \boldsymbol{\Sigma}_{22}^{-1} \boldsymbol{\Sigma}_{21} \boldsymbol{\Sigma}_{11}^{-1/2} \mathbf{t}_j$$

$$= \lambda_j^{-1} \mathbf{t}_\mathbf{i}' \mathbf{T} \mathbf{T}' \mathbf{t}_j = \begin{cases} 0 & i \neq j \\ \lambda_i & i = j \end{cases}$$

重庆工商大学
CHONGQING TECHNOLOGY AND BUSINESS UNIVERSITY

$$i = 1, 2, \cdots, \min(p_1, p_2)$$

同对相关系数为ρ_i，不同对则为零。

4、原始变量与典型变量之间的相关系数

设原始变量相关系数矩阵 $\mathbf{R}=\begin{bmatrix}\mathbf{R}_{11} & \mathbf{R}_{12}\\ \mathbf{R}_{21} & \mathbf{R}_{22}\end{bmatrix}$

x典型变量系数矩阵

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

重庆工商大学 CHONGQING TECHNOLOGY AND BUSINESS UNIVERSITY

y典型变量系数矩阵

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

重慶工商大學
CHONGQING TECHNOLOGY AND BUSINESS UNIVERSITY

X组变量与典型变量之间的关系

$$\operatorname{cov}(x_i, u_j)$$

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

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

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

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

重庆工商大学 CHONGQING TECHNOLOGY AND BUSINESS UNIVERSITY

$$\operatorname{cov}(x_i, v_j)$$

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

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

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

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

重庆工商大学
CHONGQING TECHNOLOGY AND BUSINESS UNIVERSITY

Y组变量与典型变量之间的关系

$$\mathrm{cov}(y_i, u_j)$$

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

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

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

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

重慶工商大學
CHONGQING TECHNOLOGY AND BUSINESS UNIVERSITY

$$\operatorname{cov}(y_i, v_j)$$

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

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

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

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

典型变量的结构，即变量间的相关系数		
	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

重慶工商大學
CHONGQING TECHNOLOGY AND BUSINESS UNIVERSITY

典型变量的结构，即变量间的相关系数		
	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u1的相关系数分别为0.9866和0.8872，可以看出u1可以作为消费特性的指标，第一对典型变量中v1与Y2之间的相关系数为0.9822，可见典型变量v1主要代表了了家庭收入， u1和 v1的相关系数为0.6879，这就说明家庭的消费与一个家庭的收入之间其关系是很密切的；

重慶工商大學
CHONGQING TECHNOLOGY AND BUSINESS UNIVERSITY

第二对典型变量中u2与x2的相关系数为0.4614，可以看出u2可以作为文化消费特性的指标，第二对典型变量中v2与Y1和Y3之间的分别相关系数为0.8464和0.3013，可见典型变量v2主要代表了家庭成员的年龄特征和教育程度， u2和 v2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有关。

5、各组原始变量被典型变量所解释的方差

X组原始变量被u_i解释的方差比例

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

X组原始变量被v_i解释的方差比例

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

y组原始变量被u_i解释的方差比例

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

y组原始变量被v_i解释的方差比例

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

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

典型变量的结构		
	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

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

$$m_{u_1}=(\rho^2_{u_1,x_1}+\rho^2_{u_1,x_2})/2=(0.9866^2+0.8872^2)/2=0.8803$$

$$m_{u_2}=(\rho^2_{u_2,x_1}+\rho^2_{u_2,x_2})/2=(0.1632^2+0.4614^2)/2=0.1198$$

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

$$n_{v_1} = (\rho^2_{v_1,y_1} + \rho^2_{v_1,y_2} + \rho^2_{v_1,y_q})/3$$

$$= (0.4211^2 + 0.9822^2 + 0.5145^2)/3 = 0.4689$$

$$n_{v_2} = (\rho^2_{v_2,y_1} + \rho^2_{v_2,y_2} + \rho^2_{v_2,y_q})/3$$

$$= (0.8464^2 + 0.1101^2 + 0.3013^2)/3 = 0.2731$$

典型变量的结构		
	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

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

$$m_{v_1} = (\rho^2_{v_1,x_1} + \rho^2_{v_1,x_2}) / 2 = (0.6787^2 + 0.6104^2) / 2 = 0.4166$$

$$m_{v_2} = (\rho^2_{v_2,x_1} + \rho^2_{v_2,x_2}) / 2 = (0.0305^2 + 0.0862^2) / 2 = 0.0042$$

重庆工商大学
CHONGQING TECHNOLOGY AND BUSINESS UNIVERSITY

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

$$n_{u_1} = (\rho^2_{u_1,y_1} + \rho^2_{u_1,y_2} + \rho^2_{u_1,y_q})/3$$

$$= (0.2897^2 + 0.6757^2 + 0.3539^2)/3 = 0.2219$$

$$n_{u_2} = (\rho^2_{u_2,y_1} + \rho^2_{u_2,y_2} + \rho^2_{u_2,y_q})/3$$

$$= (0.1582^2 + 0.0206^2 + 0.0563^2)/3 = 0.0095$$

重慶工商大學 CHONGQING TECHNOLOGY AND BUSINESS UNIVERSITY

6.简单相关系数、相关系数和复相关系数之间的关系

在p=1和q = 1时，X和Y之间的典型相关就是他们之间的简单相关，也就是他们的复相关。

重庆工商大学
CHONGQING TECHNOLOGY AND BUSINESS UNIVERSITY

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。由于估计中抽样误差的存在，所以估计以后还需要进行有关的假设检验。

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

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

$$
\mathbf{Z} = \begin{bmatrix}
x_{11}-\bar{x}_1 & \cdots & x_{1p}-\bar{x}_p & y_{11}-\bar{y}_1 & \cdots & y_{1q}-\bar{y}_q \\
x_{21}-\bar{x}_1 & \cdots & x_{2p}-\bar{x}_p & y_{21}-\bar{y}_1 & \cdots & y_{2q}-\bar{y}_q \\
x_{31}-\bar{x}_1 & \cdots & x_{2p}-\bar{x}_p & y_{31}-\bar{y}_1 & \cdots & y_{3q}-\bar{y}_q \\
x_{41}-\bar{x}_1 & \cdots & x_{4p}-\bar{x}_p & y_{41}-\bar{y}_1 & \cdots & y_{4q}-\bar{y}_q \\
\vdots & & \vdots & \vdots & & \vdots \\
x_{n1}-\bar{x}_1 & \cdots & x_{np}-\bar{x}_p & y_{n1}-\bar{y}_1 & \cdots & y_{nq}-\bar{y}_q
\end{bmatrix}
$$

样本的协方差：$\hat{\Sigma} = \dfrac{1}{n-1}\mathbf{Z}'\mathbf{Z} = \dfrac{1}{n-1}\begin{pmatrix} S_{xx} & S_{xy} \\ S_{yx} & S_{yy} \end{pmatrix}$

2、计算特征根和特征向量

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

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

求M_1和 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。

六、典型相关系数的检验

典型相关分析是否恰当，应该取决于两组原变量之间是否相关，如果两组变量之间毫无相关性而言，则不应该作典型相关分析。用样本来估计总体的典型相关系数是否有误，需要进行检验。

重庆工商大学
CHONGQING TECHNOLOGY AND BUSINESS UNIVERSITY

（一）整体检验

$$(H_0:\Sigma_{xy}=0;H_1:\Sigma_{xy}\neq 0)$$

$H_0:\rho_1=\cdots=\rho_r=0$,即典型相关系数均为零;

$H_1:\rho_i(i=1,2,\cdots,r)$中至少$\rho_1$不为零

检验的统计量 $\Lambda_0=\dfrac{|S|}{|S_{xx}||S_{yy}|}$

因为 $$\mathbf{S}=\begin{bmatrix}\mathbf{S}_{xx} & \mathbf{S}_{xy}\\ \mathbf{S}_{yx} & \mathbf{S}_{yy}\end{bmatrix}$$

又 $\because$ $$\begin{bmatrix}\mathbf{I} & \mathbf{0}\\ -\mathbf{S}_{yx}\mathbf{S}_{xx}^{-1} & \mathbf{I}\end{bmatrix}\begin{bmatrix}\mathbf{S}_{xx} & \mathbf{S}_{xy}\\ \mathbf{S}_{yx} & \mathbf{S}_{yy}\end{bmatrix}\begin{bmatrix}\mathbf{I} & -\mathbf{S}_{xx}^{-1}\mathbf{S}_{xy}\\ \mathbf{0} & \mathbf{I}\end{bmatrix}$$

$$=\begin{bmatrix}\mathbf{S}_{xx} & \mathbf{0}\\ \mathbf{0} & \mathbf{S}_{yy}-\mathbf{S}_{yx}\mathbf{S}_{xx}^{-1}\mathbf{S}_{xy}\end{bmatrix}$$

所以，两边同时求行列式，有

$$\begin{vmatrix} \mathbf{I} & \mathbf{0} \\ -\mathbf{S}_{yx}\mathbf{S}_{xx}^{-1} & \mathbf{I} \end{vmatrix} \begin{vmatrix} \mathbf{S}_{xx} & \mathbf{S}_{xy} \\ \mathbf{S}_{yx} & \mathbf{S}_{yy} \end{vmatrix} \begin{vmatrix} \mathbf{I} & -\mathbf{S}_{xx}^{-1}\mathbf{S}_{xy} \\ \mathbf{0} & \mathbf{I} \end{vmatrix} = \begin{vmatrix} \mathbf{S}_{xx} & \mathbf{S}_{xy} \\ \mathbf{S}_{yx} & \mathbf{S}_{yy} \end{vmatrix}$$

$$|\mathbf{S}| = \begin{vmatrix} \mathbf{S_{xx}} & \mathbf{S_{xy}} \\ \mathbf{S_{yx}} & \mathbf{S_{yy}} \end{vmatrix} = \left|\mathbf{S}_{xx}\right|\left|\mathbf{S_{yy}} - \mathbf{S_{xy}}\mathbf{S_{xx}^{-1}}\mathbf{S_{yx}}\right|$$

$$= \left|\mathbf{S_{yy}}\right|\left|\mathbf{S_{xx}}\right|\left|\mathbf{I} - \mathbf{S}_{yy}^{-1}\mathbf{S_{xy}}\mathbf{S}_{xx}^{-1}\mathbf{S_{yx}}\right|$$

$$\Lambda_0 = \frac{|\mathbf{S}|}{|\mathbf{S_{xx}}||\mathbf{S_{yy}}|} = \left|\mathbf{I} - \mathbf{S}_{yy}^{-1}\mathbf{S_{xy}}\mathbf{S}_{xx}^{-1}\mathbf{S_{yx}}\right| = \left|\mathbf{I} - \hat{\mathbf{M}}\right|$$

重慶工商大學
CHONGQING TECHNOLOGY AND BUSINESS UNIVERSITY

由于 $\left|\hat{\lambda}\mathbf{I}-\hat{\mathbf{M}}\right| = \left|\hat{\lambda}\mathbf{I}-\mathbf{I}+\mathbf{I}-\hat{\mathbf{M}}\right| = -\left|(1-\hat{\lambda})\mathbf{I}-(\mathbf{I}-\hat{\mathbf{M}})\right|$

所以若M的特征根为λ_i，则(I-M)的特征根为(1-λ_i)。

根据矩阵行列式与特征根的关系，可得：

$$\Lambda_0 = \frac{|\mathbf{S}|}{|\mathbf{S_{xx}}||\mathbf{S_{yy}}|} = \left|\mathbf{I}-\mathbf{S}_{yy}^{-1}\mathbf{S_{xy}}\mathbf{S}_{xx}^{-1}\mathbf{S_{yx}}\right| = \left|\mathbf{I}-\hat{\mathbf{M}}\right|$$

$$= (1-\lambda_1)(1-\lambda_2)\cdots(1-\lambda_p) = \prod_{i=1}^{p}(1-\lambda_i)$$

Λ_0小，则支持备择假设。

重慶工商大學 CHONGQING TECHNOLOGY AND BUSINESS UNIVERSITY

在原假设为真的情况下，检验的统计量

$$Q=-[(n-1)-(p+q+1)/2]\ln\Lambda_0$$

近似服从自由度为pq的χ^2分布。在给定的显著性水平α下，如果$\chi^2\geq\chi^2$（pq)，则拒绝原假设，认为至少第一对典型变量之间的相关性显著。

依此类推，再检验下一对典型变量之间的相关性。直至相关性不显著为止。对两组变量x和y进行典型相关分析，采用的也是一种降维技术。我们希望使用尽可能少的典型变量对数，为此需要对一些较小的典型相关系数是否为零进行假设检验。H_0经检验被拒绝，则应进一步检验假设。

（二）部分总体典型相关系数为零的检验

$$H_0: P_2 = \dots = P_r = 0$$

$H_1: P_2$，P_3，P_r至少有一个不为零。若原假设H_0被接受，则认为只有第二对典型变量是有用的；若原假设H_0被拒绝，则认为第二对典型变量也是有用的，并进一步检验假设

重庆工商大学
CHONGQING TECHNOLOGY AND BUSINESS UNIVERSITY

$H_0: P_3 = \ldots = P_r = 0$

$H_1: P_3$，…，P_r至少有一个不为零。

如此进行下去. 直至对某个k，

$H_0: P_{(k+1)} = \ldots = P_M = 0$

$H_1: P_{(k+1)}, \ldots, P_m$至少有一个不为零

检验的统计量 $$\Lambda_k=\prod_{i=k+1}^{r}(1-\rho_i^2)=\prod_{i=k+1}^{r}(1-\lambda_i)$$

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

近似服从自由度为(p-k)(q-k)的χ^2分布。在给定的显著性水平α下，如果$\chi^2 \geq \chi^2$[(p-k)(q-k)]，则拒绝原假设，认为至少第k+1对典型变量之间的相关性显著。

H0：当前和后面的典型相关系数均为零

H1：至少当前的典型相关系数为零

	Likelihood Ratio	Approx F	Num DF	Den DF	Pr > F
1	0.50833498	1341.234	6	19990	0.0001
2	0.96508130	180.838	2	9996	0.0001

可见，前面两对典型变量的相关性是很强的。

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

重慶工商大學
CHONGQING TECHNOLOGY AND BUSINESS UNIVERSITY

	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

重慶工商大學
CHONGQING TECHNOLOGY AND BUSINESS UNIVERSITY

当前和后面的典型相关系数均为零的检验

	Likelihood Ratio	Approx F	Num DF	Den DF	Pr>F
1	0.63988477	134.4237	35	42018.15	0.0001
2	0.92280941	33.8242	24	34848.67	0.0001
3	0.97743541	15.2634	15	27578.39	0.0001
4	0.99152030	10.6579	8	19982	0.0001
5	0.99672015	10.9600	3	9992	0.0001

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

重慶工商大學
CHONGQING TECHNOLOGY AND BUSINESS UNIVERSITY

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

重慶工商大學
CHONGQING TECHNOLOGY AND BUSINESS UNIVERSITY

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成员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。

重慶工商大學
CHONGQING TECHNOLOGY AND BUSINESS UNIVERSITY

Canonical Redundancy Analysis

Raw Variance of the 'VAR' Variables Explained by

	Their Own Canonical Variables Proportion	Their Own Canonical Variables Cumulative Proportion	The Opposite Canonical Variables Proportion	The Opposite Canonical Variables 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
4	**0.1223**	**0.9081**	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 Proportion	Their Own Canonical Variables Cumulative Proportion	The Opposite Canonical Variables Proportion	The Opposite Canonical Variables 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
4	**0.1289**	**0.6972**	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：

$$m_{u_1} = \frac{1}{5}(0.83^2 + 0.74^2 + \cdots + 0.85^2) = 0.5818$$

$$n_{v_1} = \frac{1}{7}(0.75^2 + 0.65^2 + \cdots + 0.50^2) = 0.3721$$

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%

我国居民消费构成及主要影响因素

典型相关分析的基本思想：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，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，第二对本身具有最大相关性。如此下去，直至两组变量的相关性被提取完为止。本例想利用我国1999年城镇居民的家庭收入来源和消费性支出的数据了解我国居民消费构成及主要影响因素分析所用的数据来自：《中国统计年鉴》2000。

收入指标：X1——可支配收入
X2——实际收入
X3——国有单位职工收入
X4——集体单位职工收入
X5——其他经济类型职工收入，
X6——转移收入

支出指标：Y1——消费性支出
Y2——食品
Y3——衣着
Y4——交通和通讯
Y5——医疗和保健
Y6——娱乐、教育、文化服务
Y7——居住

序号	典型相关系数	典型变量
1	0.990174	U_1 =0.9989X1+-0.0595X2+0.0776X3 +0.0489X4-0.0931X5+0.0074X6 V_1=1.3263Y1-0.0270Y2-0.0005Y3-0.0769Y4 -0.0717Y5-0.2031Y6-0.0219Y
2	0.868704	U_2 =-4.8668X1+0.1264X2+1.9585X3 +0.3299X4+1.4095X5+2.6453X6 V_2 =-4.4920Y1+2.5421Y2+1.2480Y3-0.4621Y4 +1.0443Y5+0.8610Y6+0.0586Y7

重庆工商大学
CHONGQING TECHNOLOGY AND BUSINESS UNIVERSITY

由累计贡献率得知,第一组和第二组变量的累计贡献率已达到了97.56%,而且,这两组的系数和方差与其他组相比要大得多.即只需要前两组变量就已经可以解释全部信息的97.56%.

在第一对典型变量中，U_1 主要受可支配收入的影响，V_1 主要受消费性支出的影响；可见实际收入对消费支出的影响远小于可支配收入的影响。居民消费主要依据其可支配收入而定。

第二对典型变量中，U_2 主要受国有单位职工收入、其他经济类型职工收入和转移收入的影响，V_2 主要受食品、衣着、医疗和保健的影响。

可见我国集体单位的职工收入还不能够与国有甚至是其他经济类型的单位这职工收入相比，也从一个侧面放反映了集体单位规模等方面的现状。再有就是我国居民食品和衣着方面的支出仍占了总支出的大部分，反映了我国居民总体收入水平还不够高；其次，医疗保健支出的比例比较大是可喜的，说明我国居民已经可以把部分精力放在了自己身体的调养上来，全国居民的总体健康状况在上升之中。让我们担忧的是在教育方面的支出所占比例太小，不符合现今世界发展对教育程度的要求 。科技是第一生产力，如何提高国民的科技文化知识水平是当今的一大重点。在当代激烈的竞争中，没有知识的支撑是不行的。

重慶工商大學
CHONGQING TECHNOLOGY AND BUSINESS UNIVERSITY

重慶工商大学

CHONGQING TECHNOLOGY AND BUSINESS UNIVERSITY