

入学成绩对大学生基础课成绩的影响分析*

陆 媛

(盐城工学院 基础教学部, 江苏 盐城 224003)

摘 要 通过对某校2001级100名本科(理工)学生的入学成绩和大学一年级基础课成绩进行统计,运用典型相关分析方法研究这两种成绩,分析了二者之间的相关性,得出入学成绩中数学和物理成绩对高等数学、普通物理的影响较大。

关键词 典型相关变量 相关矩阵 影响分析

中图分类号: O212.1

文献标识码: A

文章编号: 1671-5322(2005)04-0050-03

典型相关分析是多元统计中的一种重要的统计方法,其作用在于研究两个因子集团之间的关系。它的基本思想是在两组变量中选择若干个有代表性的综合变量,每一个综合变量都是这组变量的线性组合,而且综合变量之间是相关的,这种综合变量称为典型相关变量。再通过对这一组典型相关变量相关关系的研究来代替原来两组变量之间相互关系的研究。在实践中,只需要着重研究相关系数较大的那几对典型相关变量。

1 引理和方法

设两组变量分别有 m, n 个变量,且 $p_1 = m, p_2 = n$, 假设两组变量分别为:

$X_1 = [x_1, x_2, \dots, x_n]^T, X_2 = [x_{m+1}, x_{m+2}, \dots, x_{m+n}]^T$
变量之间的相关性只与它们的协方差有关,设协方差矩阵为 C , 根据两组向量的维数把 C 分成4

块 $C = \begin{bmatrix} C_{11} & C_{12} \\ C_{21} & C_{22} \end{bmatrix}$, 其中 C_{ij} 是 $(p_i \times p_j)$ 阶矩阵, 由

协方差矩阵的性质可知 C 为对称正定矩阵。

研究两组变量或者两个随机变量之间的相关性,等价于研究 X_1 与 X_2 之间的相关性。当 $C_{12} = 0$ 时, $\text{Cov}[X_1, X_2] = 0$, X_1 与 X_2 不相关, 讨论 X_1 与 X_2 之间的典型相关性就没必要以下讨论是在假设 $C_{12} \neq 0$ 时。设 X_1 与 X_2 的线性组合分别为 $U = \alpha^T X_1, V = \beta^T X_2$, α, β 分别是 m, n 维列向量。

推论: U 与 V 之间的相关系数在线性变换下是不变的。

证明: 对于任意实数 $c \neq 0, d \neq 0$, 且 $c, d, e, f \in R$, 则有:

$$\begin{aligned} \rho(cU + d, eV + f) &= c\text{Cov}(U, V) / \sqrt{c^2 D(U)} \sqrt{e^2 D(V)} \\ &= \text{Cov}(U, V) / \sqrt{D(U)D(V)} = \rho(U, V) \end{aligned}$$

其中 ρ 是向量的相关系数, $D(Y)$ 为向量 Y 的方差。选用适当的 m, n 维列向量 a, b 使 $D(U) = 1, D(V) = 1$,

$$\text{引理 } \rho(U, V) = \text{Cov}(U, V) = \alpha^T C_{12} b^{[3]}$$

选取合适的向量 α, β 使得在 $D(U) = 1, D(V) = 1$ 的情况下, 相关系数达到最大。用拉格朗日乘数法来求得, 设拉格朗日乘数 r 及 y , 建立目标函数

$$Q(a, b) = \alpha^T C_{12} b - \frac{r(\alpha^T C_{11} a - 1)}{2} - y \frac{\beta^T C_{22} b - 1}{2}$$

此函数达到最大值的必要条件为

$$\frac{\partial Q}{\partial a} = C_{12} b - r C_{11} a = 0, \frac{\partial Q}{\partial b} = C_{21} a - y C_{22} b = 0$$

可推出 $r = y = \alpha^T C_{21} b = \rho(U, V)$, 所以 r 为 U, V 的相关系数。经过推理可得到:

$$(C_{12} C_{22}^{-1} C_{21} - r^2 C_{11}) \alpha = (C_{21} C_{11}^{-1} C_{12} - r^2 C_{22}) \beta = 0$$

选取矩阵 D_1 , 使得 $D_1^{-1} = D_1^{-12} = D_1^{-12}$ 记 $T =$

$$C_{11}^{-\frac{1}{2}} C_{12} C_{22}^{-\frac{1}{2}}, \alpha = C_{11}^{1/2} a, \beta = C_{22}^{1/2} b \text{ 从而 } \alpha^T \alpha$$

* 收稿日期: 2005-08-07

作者简介: 陆媛(1976-), 江苏滨海人, 盐城工学院基础教学部助教, 主要研究方向为高等数学教学及理论。

$=1 \beta^T \beta = 1$ 经过简单推理可得到 $(TT^T - r^2 E_m)$
 $\alpha = O(1) (TT^T - r^2 E_n) \beta = O(2)$ 其中 E 为单位
 矩阵, O 为零矩阵, 则 α, β 有非零解的必要条件
 为: $|TT^T - r^2 E_m| = 0, |T^T T - r^2 E_n| = 0$

设 $m \geq n$, 其中 TT^T 是 m 阶非负定阵, 具有 m
 个非零特征值, $T^T T$ 是 n 阶非负定阵, 具有 n 个非
 零特征值。可证明 TT^T 和 $T^T T$ 具有相同的非零特
 征值^[3]

假设 $T^T T$ 的非零特征值为 $r_1^2 \geq r_2^2 \geq \dots \geq r_n^2 > 0$,
 现在由 (1) (2) 分别求得特征向量 α_i, β_i 从而得
 $a_i = C_{11}^{-1/2} \alpha_i, b_i = C_{22}^{-1/2} \beta_i$ 并建立 n 对线性组合
 $u_i = a_i^T X_1, v_i = b_i^T X_2$, 此时 $(u_i, v_i) = r_i$, 以上 $i = 1,$
 $2, \dots, n$ 。可根据需要选取具体的前几对典型相关
 变量。

2 主要应用成果

2.1 本文研究的问题

$$S = (s_{ij})_{7 \times 7} = \begin{pmatrix} 251.704 & 83.781 & 135.596 & 134.458 & 175.703 & 54.299 & 100.627 \\ 83.781 & 233.840 & 116.299 & 90.719 & 51.044 & 126.453 & 69.722 \\ 135.596 & 116.299 & 231.530 & 119.202 & 99.830 & 63.720 & 166.497 \\ 134.458 & 90.719 & 119.202 & 219.046 & 100.618 & 51.411 & 79.557 \\ 175.703 & 51.044 & 99.830 & 100.618 & 171.438 & 43.198 & 88.369 \\ 54.299 & 126.453 & 63.720 & 51.411 & 43.198 & 99.568 & 53.487 \\ 100.627 & 69.722 & 166.497 & 79.557 & 88.369 & 53.487 & 195.057 \end{pmatrix}$$

和相关矩阵

$$R = (r_{ij})_{7 \times 7} = \begin{pmatrix} 1.000 & 0 & 0.345 & 3 & 0.561 & 7 & 0.572 & 6 & 0.845 & 8 & 0.343 & 0 & 0.454 & 1 \\ 0.345 & 3 & 1.000 & 0 & 0.499 & 4 & 0.400 & 8 & 0.254 & 9 & 0.828 & 7 & 0.326 & 5 \\ 0.561 & 7 & 0.499 & 4 & 1.000 & 0 & 0.529 & 3 & 0.501 & 1 & 0.419 & 7 & 0.783 & 5 \\ 0.572 & 6 & 0.400 & 8 & 0.529 & 3 & 1.000 & 0 & 0.519 & 2 & 0.348 & 1 & 0.384 & 9 \\ 0.845 & 8 & 0.254 & 9 & 0.501 & 1 & 0.519 & 2 & 1.000 & 0 & 0.330 & 6 & 0.483 & 2 \\ 0.343 & 0 & 0.828 & 7 & 0.419 & 7 & 0.348 & 1 & 0.330 & 6 & 1.000 & 0 & 0.383 & 8 \\ 0.454 & 1 & 0.326 & 5 & 0.783 & 5 & 0.384 & 9 & 0.483 & 2 & 0.383 & 8 & 1.000 & 0 \end{pmatrix}$$

S 和 R 为对称正定矩阵, 这里称为相关系数。记

$$A_1 = \text{diag}(\sqrt{s_{ij}}) = \text{diag}(15.862 \ 5 \ 15.291 \ 8 \ 15.216 \ 1 \ 14.800 \ 2),$$

$$A_2 = \text{diag}(\sqrt{s_{ij}}) = \text{diag}(13.093 \ 4 \ 90.978 \ 4 \ 13.966 \ 3),$$

$$A = \text{diag}(\sqrt{s_{ij}}) = \begin{pmatrix} A_1 & 0 \\ 0 & A_2 \end{pmatrix} (i = 1 \ 2 \ 3 \ 4 \ j = 5 \ 6 \ 7)$$

则有 $S = ARA$ 。可把 S 当作协方差矩阵的近似。
 根据典型相关分析的方法, 先把 S, R 分成 4 个部

$$\text{分}, R = \begin{pmatrix} R_{11} & R_{12} \\ R_{21} & R_{22} \end{pmatrix}, S = \begin{pmatrix} S_{11} & S_{12} \\ S_{21} & S_{22} \end{pmatrix} \text{ 其中 } R_{11}, S_{11}$$

为 4 阶方阵, R_{22}, S_{22} 为 3 阶方阵, 且有 $S_{11} = A_1 R_{11} A_1, S_{12} = A_1 R_{12} A_2, S_{21} = A_2 R_{21} A_1, S_{22} = A_2 R_{22} A_2$, 则求

本文所有成绩数据来源于某校 2001 级计算
 机、建工、材料等专业部分学生成绩, 采用随机抽
 样的方法得到。分别把入学成绩和大一基础课成
 绩当作两个因子集团, 分别记为 X_1, X_2 其中 $x_1,$
 x_2, x_3, x_4 分别表示入学成绩中的数学、英语、物
 理、化学, x_5, x_6, x_7 分别表示大学一年级高等数
 学、英语、普通物理。 $X_1 = [x_1 \ x_2 \ x_3 \ x_4]^T, X_2 =$
 $[x_5 \ x_6 \ x_7]^T$ 这里令 $m = 4, n = 3$ 。建立 X_1 与 X_2
 的线性组合 $U = \alpha^T X_1, V = \beta^T X_2$ 这里 α, β 分别是
 4, 3 维列向量, 本文取了容量为 100 的样本。设
 $N = 100$ 根据公式 $\bar{x}_i = \sum_{i=1}^N x_{ii} / N \quad i = 1 \ 2 \ \dots \ 7 \quad t$
 $= 1 \ 2 \ \dots \ N$ 算出均值向量 $\bar{X} = [x_1 \ x_2 \ \dots \ x_n]$
 $[110.42 \ 104.14 \ 104.51 \ 96.29 \ 71.04 \ 62.55 \ 69.77]$ 再根据公式 $s_{ij} = \sum_{i=1}^N (x_{ii} - \bar{x}_i)(x_{ii} - \bar{x}_j) / N,$
 $r_{ij} = s_{ij} / \sqrt{s_{ii} s_{jj}} \quad i, j = 1 \ 2 \ \dots \ 7 \quad t = 1 \ 2 \ \dots \ N$ 得:

特征值 λ 的方程为 $|S_{21} S_{11}^{-1} S_{12} - \lambda^2 S_{22}| = 0$

(3) 即 $|A_2| \cdot |R_{21} R_{11}^{-1} R_{12} - \lambda^2 R_{22}| \cdot |A_2| =$
 0 , 又因为 $|A_2| \neq 0$, 所以有: $|R_{21} R_{11}^{-1} R_{12} -$
 $\lambda^2 R_{22}| = 0$ (4) (3) 和 (4) 的特征根相同, 用
 $T^T T - \lambda^2 I_n = 0$ 求得特征值^[3]

$$\lambda_1^2 = 0.742 \ 6 \ \lambda_2^2 = 0.200 \ 8 \ \lambda_3^2 = 0.153 \ 5,$$

$$\therefore \lambda_1 = 0.861 \ 7 \ \lambda_2 = 0.448 \ 1 \ \lambda_3 = 0.391 \ 8$$

同时求得与之对应的特征向量为:

$$\begin{cases} y_1 = [0.818 \ 4 \ 0.308 \ 0 \ 0.485 \ 2]^T \\ y_2 = [-0.282 \ 2 \ 0.950 \ 8 \ -0.127 \ 6]^T \\ y_3 = [-0.500 \ 6 \ -0.032 \ 5 \ 0.865 \ 1]^T \end{cases}$$

所以：

$$\begin{cases} A_2\beta_1=[0.757\ 6\ 0.146\ 6\ 0.291\ 4]^T \\ A_2\beta_2=[-0.402\ 9\ 1.084\ 0\ -0.238\ 5]^T, \\ A_2\beta_3=[-0.782\ 1\ -0.123\ 0\ 1.125\ 4]^T \end{cases}$$

$$\begin{cases} A_1\alpha_1=[0.740\ 3\ 0.046\ 5\ 0.319\ 8\ 0.03\ 4]^T \\ A_1\alpha_2=[0.298\ 7\ -1.143\ 0\ 0.305\ 2\ 0.032\ 5]^T \\ A_1\alpha_3=[-0.840\ 2\ -0.192\ 8\ 1.260\ 5\ -0.131\ 2]^T \\ \alpha_1=[0.046\ 7\ 0.003\ 0\ 0.021\ 0\ 0.002\ 3]^T \\ \alpha_2=[0.018\ 8\ -0.070\ 4\ 0.020\ 1\ 0.002\ 2]^T \\ \alpha_3=[-0.053\ 0\ -0.012\ 3\ 0.082\ 8\ -0.008\ 9]^T \end{cases}$$

和

$$\begin{cases} \beta_1=[0.057\ 9\ 0.014\ 7\ 0.020\ 9]^T \\ \beta_2=[-0.030\ 8\ 0.108\ 6\ -0.017\ 1]^T \\ \beta_3=[-0.059\ 7\ -0.012\ 3\ 0.080\ 6]^T \end{cases}$$

所以得到三组典型相关变量：

$$\begin{cases} U_1=0.046\ 7x_1+0.003\ 0x_2+0.021\ 0x_3+0.002\ 3x_4; \\ V_1=0.057\ 9x_5+0.014\ 7x_6+0.020\ 9x_7 \\ U_2=0.018\ 8x_1-0.074\ 7x_2+0.021\ 0x_3+0.002\ 2x_4; \\ V_2=-0.030\ 8x_5+0.108\ 6x_6-0.017\ 1x_7 \\ U_3=-0.053\ 0x_1-0.012\ 3x_2+0.082\ 8x_3-0.008\ 9x_4 \\ V_3=-0.057\ 9x_5-0.012\ 3x_6+0.080\ 6x_7 \end{cases}$$

2.2 简单的显著性检验

运用 Bartett 法检验典型相关系数的显著性。

取 $\alpha=0.01$ 则根据 Bartett 检验法：

$$\alpha_i=(N-1)-\frac{(m+n+1)}{2}, \text{自由度 } f_i=(m-i) \times (n-i);$$

且当 $w_i=I(1-\lambda_j^2)$ 若 $-\alpha_i \log(w_i) \geq x_j^2(\alpha)$ 时,说

参考文献：

[1] 杨维权,刘兰亭,林鸿洲.多元统计分析[M].第2版.北京:高等教育出版社,1984.

明拒绝假设中的即第 i 组的典型相关系数是显著的,其中 $x_j^2(\alpha)$ 可查临界值得到。

表 1 计算结果

Table 1 Calculation result

λ	w	$-\alpha \log(w)$	f	$x_j^2(\alpha)$	结论
0.861 7	0.174 1	72.225 4	12	26.217	拒绝
0.448 1	0.676 5	15.953 5	6	16.812	接受
0.391 8	0.846 5	6.730 2	2	9.210	接受

由表可知,第一组典型相关系数是显著的,其它组都不显著。所以,典型相关只有 U_1, V_1 一组。

3 结论

根据本文的分析结果,入学成绩与大一基础课成绩之间有较强的相关关系。通常,入学成绩好,说明中学的基础好,这样入学后对基础课的学习有促进作用,经显著性检验结果表明,只需用 (U_1, V_1) 来分析 X_1 与 X_2 之间的相关关系,且 $\rho(U_1, V_1)=0.861\ 7$, 即 U_1, V_1 之间有高度的正相关关系。在 U_1 中 x_1 (入学数学成绩)的权系数最大,其次是 x_3 (物理)的系数,可见 U_1 对 V_1 起促进作用主要是依赖于 x_1 数学和 x_3 物理; V_1 中权系数最大的是 x_5 (高等数学),其次是 x_7 (普通物理)。说明 U_1 对 V_1 的影响最大的是 x_5 高等数学,其次是 x_7 普通物理,然后是外语在其中的分量也不小。所以入学成绩中数学和物理对一年级基础课中高等数学和普通物理的影响较大,其次是英语成绩的影响。

The Analysis of Relation Between Entrance Examination Scores and Basic Subjects Scores of Freshmen

LU Yuan

(Department of Foundamental Sciences Teaching, Yancheng Institute of Technology, Jiangsu Yancheng 224003, China)

Abstract The author applies canonical analysis to study the relation between entrance examination scores and basic subject scores of freshmen. Concludes that entrance examination mathematics score and the physics score have more important effect on the basic subjects advanced mathematics and ordinary physics, especially on the advanced mathematics and proposes three pieces of advice to our educational reform.

Keywords Canonical correlation variable, correlational matrix, effect analysis