第44卷 第1期 气 象 学 报 Vol. 44, No. 1
1986 年 2 月 ACTA METEOROLOGICA SINICA Feb., 1986

论变量线性相关系数的相互制约及其对多元线性回归方程拟合优度的影响*

王 盘 兴

(南京气象学院)

提 要

本文提出并定量地研究了三个变量间线性相关系数的相互制约关系。在此基础上讨论了因子相关对多元线性回归方程拟合优度的影响。提出一种选因子的标准，并初步论证了它的合理性。

一、引 言

多元线性回归方程是气象科学中广泛使用的统计学方法之一。天气预报领域内回归分析的实践指出，回归方程的预报效果极大地依赖于预报因子的合理选择[1]。有效的预报因子应当与预报量间有清楚的物理联系和好的相关关系。但是，在很多场合(如长期天气预报中)，我们对前后期天气间因果联系的物理机制知之甚少，选择预报因子不得不主要地依靠统计手段。经验认为，有效的预报因子应符合下述要求：(1) 因子与预报量的相关系数绝对值较大；(2) 因子与因子间的相关系数绝对值要小。

然而，近年来的实践提出了一些新问题[2]。例如，使用某些方案建立多元回归方程时，其最终入选的因子并不一定是供选择的所有因子中最能满足上述两个要求的那些因子。一些作者[3]研究了这个问题，提出：某些情况下因子间存在一定相关时有利于提高多元线性回归方程的预报效果。统计学在其它学科中的应用也遇到类似的问题，M. 肯德尔[4]研究了这类现象，并对此作了定性分析。这就对统计学天气预报中究竟应当按何种原则选择预报因子的问题提出了与传统观念相对立的一种意见。

本文从线性代数基本理论出发，导出了变量线性相关系数间存在的制约关系。在此基础上定量地讨论了因子相关对多元线性回归方程拟合优度的影响，提出了一种选择预报因子的原则，并对它的合理性作了初步论证。

二、变量线性相关系数间的制约关系

与预报量Y具有线性相关系数r_1,r_2的预报因子X_1、X_2间的线性相关系数r_{12}之间

* 本文于1984年5月7日收到，1984年11月14日收到修改稿。

存在何种关系，是本节研究的问题。

设样本容量为 n，则 Y, X_1, X_2 可以表作实数域上 n 维线性空间 Γ 中的向量

$$Y=\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} \quad X_1=\begin{bmatrix} x_{11} \\ x_{12} \\ \vdots \\ x_{1n} \end{bmatrix} \quad X_2=\begin{bmatrix} x_{21} \\ x_{22} \\ \vdots \\ x_{2n} \end{bmatrix} \tag{1}$$

它们的标准化形式为

$$Y^0=\begin{bmatrix} y_1^0 \\ y_2^0 \\ \vdots \\ y_n^0 \end{bmatrix} \quad X_1^0=\begin{bmatrix} x_{11}^0 \\ x_{12}^0 \\ \vdots \\ x_{1n}^0 \end{bmatrix} \quad X_2^0=\begin{bmatrix} x_{21}^0 \\ x_{22}^0 \\ \vdots \\ x_{2n}^0 \end{bmatrix} \tag{2}$$

由变量相关系数及线性空间中内积定义得

$$\begin{aligned} r_1&=Y^0\cdot X_1^0 \\ r_2&=Y^0\cdot X_2^0 \\ r_{12}&=X_1^0\cdot X_2^0 \end{aligned} \tag{3}$$

将 n 维线性空间 V 分解成直交的两个子空间 U，W，其中 U 是 Y^0 所在的一维子空间，W 是 $n-1$ 维子空间。即

$$V=U\dot{+}W \tag{4}$$

式中 $\dot{+}$ 为直和记号。此时，显然有 Y^0, X_1^0, X_2^0 在子空间 U 中的分量的模

$$|Y^{0\prime}|=1,\quad |X_1^{0\prime}|=|r_1|,\quad |X_2^{0\prime}|=|r_2| \tag{5}$$

和它们在子空间 W 中的分量的模

$$|Y^{0\prime\prime}|=0,\quad |X_1^{0\prime\prime}|=\sqrt{1-r_1^2},\quad |X_2^{0\prime\prime}|=\sqrt{1-r_2^2} \tag{6}$$

设在子空间 W 中分量 $X_1^{0\prime\prime}, X_2^{0\prime\prime}$ 的交角为 φ，因此，其内积为

$$X_1^{0\prime\prime}\cdot X_2^{0\prime\prime}=\sqrt{1-r_1^2}\sqrt{1-r_2^2}\cos\varphi \tag{7}$$

注意 X_1^0, X_2^0 在子空间 U 中的分量的内积

$$X_1^{0\prime}\cdot X_2^{0\prime}=r_1r_2 \tag{8}$$

及(4)式给出的直和关系，得

$$\begin{aligned} r_{12}&=X_1^0\cdot X_2^0=(X_1^{0\prime}+X_1^{0\prime\prime})\cdot(X_2^{0\prime}+X_2^{0\prime\prime}) \\ &=X_1^{0\prime}\cdot X_2^{0\prime}+X_1^{0\prime\prime}\cdot X_2^{0\prime\prime} \end{aligned} \tag{9}$$

以(7)、(8)代入(9)式得

$$r_{12}=r_1r_2+\sqrt{1-r_1^2}\sqrt{1-r_2^2}\cos\varphi \tag{10}$$

(10)式给出了预报因子 X_1、X_2 间的相关系数 r_{12} 与相关系数 r_1, r_2 间的制约关系。其右端自变量的取值区间为

$$r_1, r_2\in[-1,1],\quad \varphi\in[0,\pi].$$

又由向量间夹角的定义知，Y^0, X_1^0, X_2^0 间的夹角 $\alpha_1, \alpha_2, \alpha_{12}$ 间有如下一一对应关系

$$\alpha_1=\widehat{Y^0X_1^0}=\arccos r_1$$

$$\alpha_2=\widehat{Y^0X_2^0}=\arccos r_2 \tag{11}$$
$$\alpha_{12}=\widehat{X_1^0X_2^0}=\arccos r_{12}$$

利用关系(11)可将 r_{12} 对 r_1,r_2,φ 的依赖关系用图 1 显示出来。

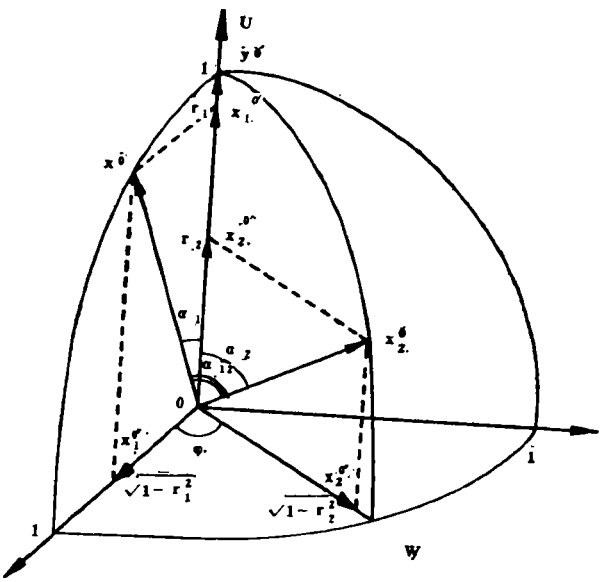


图 1 r_1,r_2,φ 对 r_{12} 的制约关系示意

对(10)式的研究得，

$\varphi=0$ 时，r_{12} 取得其极大值

$$r_{12max}=r_1r_2+\sqrt{1-r_1^2}\sqrt{1-r_2^2} \tag{12}$$

$\varphi=\pi$ 时，r_{12} 取得其极小值

$$r_{12min}=r_1r_2-\sqrt{1-r_1^2}\sqrt{1-r_2^2} \tag{13}$$

φ 在$[0,\pi]$上取值时，r_{12} 的平均值为

$$\overline{r_{12}}=\frac{1}{\pi}\int_0^{\pi}r_{12}d\varphi=r_1r_2 \tag{14}$$

由(10)式知，它也是 $X_1^{0\prime\prime},X_2^{0\prime\prime}$ 正交时的 r_{12} 值，即

$$\overline{r_{12}}=r_{12}|_{\varphi=\frac{\pi}{2}}$$

由(12)、(13)式，r_{12} 的变幅

$$\Delta r_{12}=r_{12max}-r_{12min}=2\sqrt{1-r_1^2}\sqrt{1-r_2^2} \tag{15}$$

由(12)—(15)式给出的 $r_{12max},r_{12min},\bar{r}_{12},\Delta r_{12}$ 与 r_1,r_2 的关系由图 2 给出。

图 2 表明：因子间相关的性质(r_{12} 的正或负)主要受制于因子与预报量间的相关性质，与预报量相关性质相同(r_1,r_2 同号)的两个因子间主要表现为正相关，反之，则主要表现为负相关；因子间相关的程度(r_{12} 绝对值大小)则随因子与预报量间相关程度的加强而加强；当且仅当 $r_1^2+r_2^2\leqslant 1$ 时，r_{12} 才有取得 0 值(对应因子间不相关)的可能。

三、制约关系对二元线性回归方程拟合优度的影响

给定样本容量为 n 的预报量 Y 及因子 $X_j(j=1—p)$后，可以建立一个 p 元线性回归方程

$$\hat{Y}=b_0+\sum_{j=1}^{p}b_jX_j \tag{16}$$

其中，$\hat{Y}$ 是预报量的回归估计值，b_0，b_j 是回归系数，X_j 是预报因子取值。

方程(16)的拟合优度 K 通常用它的回归平方和 u 与预报量离差平方和 s_{yy} 的比值表示。即

$$K=u/s_{yy}$$

已经证明，K 与复相关系数间存在关系

$$K=R^2 \tag{17}$$

而

$$R^2=1-D/D_0 \tag{18}$$

式中

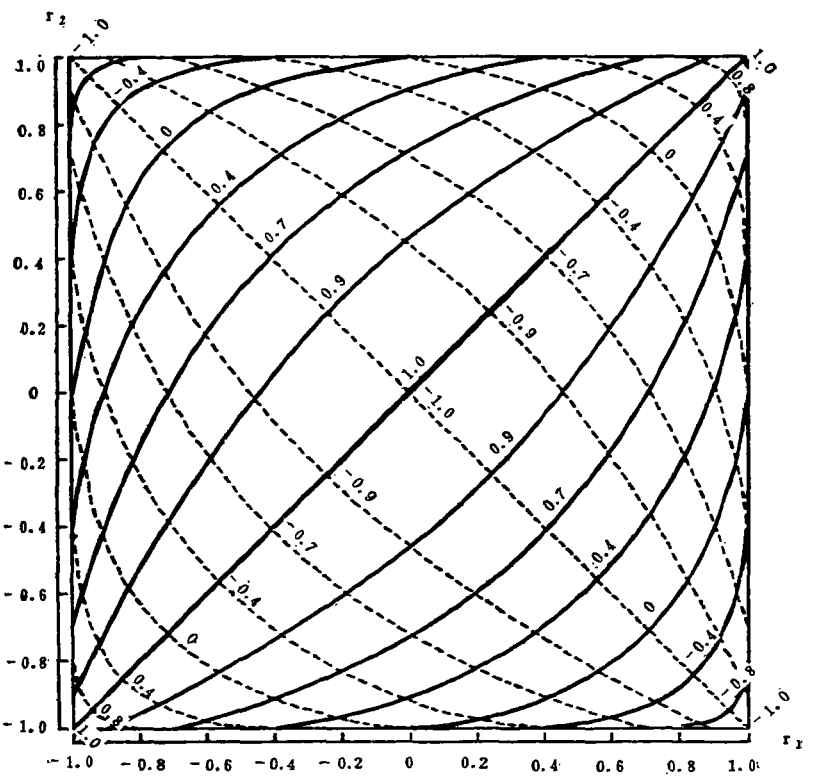


图 2a　r_{12max}(实线)和 r_{12min}(虚线)

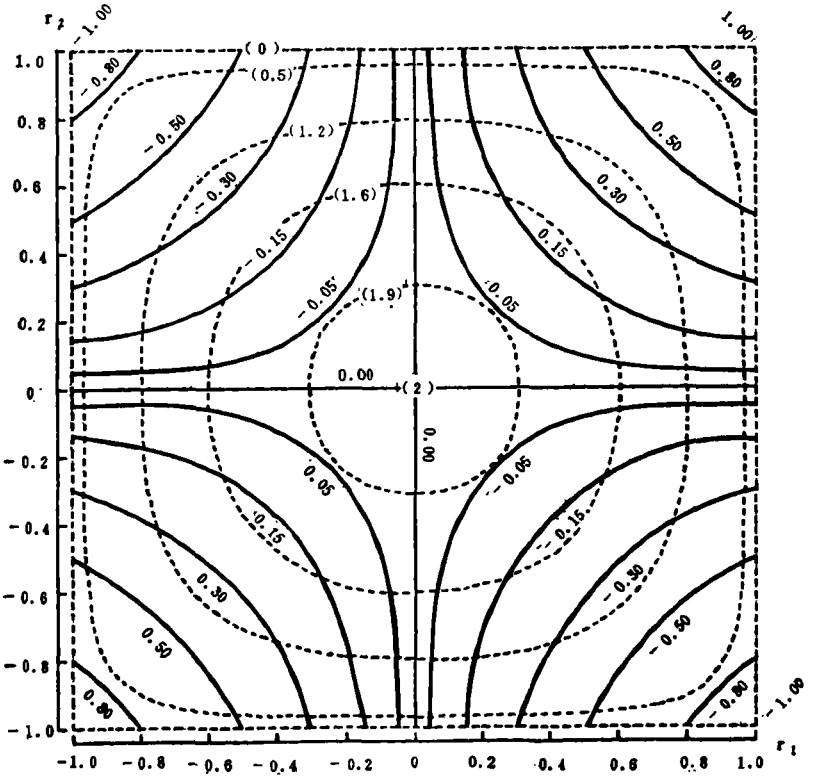


图 2b $\bar{r}_{12}$(实线)和 Δr_{12}(虚线)

$$D=\begin{vmatrix} 1 & r_1 & r_2 & \cdots & r_p \\ r_1 & 1 & r_{12} & \cdots & r_{1p} \\ r_2 & r_{21} & 1 & \cdots & r_{2p} \\ & & \cdots\cdots & & \\ r_p & r_{p1} & r_{p2} & \cdots & 1 \end{vmatrix}=D(r_i, r_{i_1 i_2}; p) \tag{19}$$

$$D_0=\begin{vmatrix} 1 & r_{12} & \cdots & r_{1p} \\ r_{21} & 1 & \cdots & r_{2p} \\ & \cdots\cdots & & \\ r_{p1} & r_{p2} & \cdots & 1 \end{vmatrix}=D_0(r_{i_1 i_2}; p) \tag{20}$$

行列式 D、D_0 的元素 r_i 是 Y 与 X_i 的相关系数，$r_{i_1 i_2}$ 是 X_{i_1} 与 X_{i_2} 的相关系数，显然 $r_{i_1 i_2}=r_{i_2 i_1}$。

另外，由(10)式得

$$r_{i_1 i_2}=r_{i_1}r_{i_2}+\sqrt{1-r_{i_1}^2}\sqrt{1-r_{i_2}^2}\cos\varphi_{i_1 i_2} \tag{21}$$

式中 $\varphi_{i_1 i_2}$ 是 $X_{i_1}^0, X_{i_2}^0$ 在子空间 W 中分量的交角。

若以 r 记 r_i 之总体，r' 记 $r_{i_1 i_2}$ 之总体，φ 记 $\varphi_{i_1 i_2}$ 之总体，则由(17)—(21)式得

$$K(r,\varphi;p)=1-\frac{D(r,\varphi;p)}{D_0(r,\varphi;p)} \tag{22}$$

(22)式表明，因子数确定后，方程(16)的拟合优度取决于预报因子与预报量相关系数的总体及子空间W中因子间交角的总体。

因为(22)式已引入了因子间相互制约关系(10)，式中r,φ有明确的定义域，用它讨论K的取值规律是方便的。我们将先讨论$p=2$的情形，侧重研究因子间相关联系对拟合优度的影响，从中引出稳定最优因子对概念，给出一种新的选因子标准。然后，利用二、三两节的结论讨论一种简化情形下多元线性回归方程拟合优度曲线的性状，考察拟合优度随因子间相关联系和因子数变化的规律。

由(22)、(19)、(20)、(10)式得二元线性回归方程拟合优度K_2的表式

$$\left.\begin{aligned} K_2&=(r_1^2+r_2^2-2r_1r_2r_{12})/(1-r_{12}^2)\\ r_{12}&=r_1r_2+\sqrt{1-r_1^2}\sqrt{1-r_2^2}\cos\varphi \end{aligned}\right\} \tag{23}$$

由(23)式知，r_1,r_2一经确定，K_2便仅是r_{12}或φ在它们所对应的定义域上的一元函数。因此，容易通过研究K_2对r_{12}或φ的导数来认识函数K_2的性质。由此得到K_2的可能极值点为：

1．对所有的r_1,r_2：

$$\left.\begin{aligned} &\varphi=0,\pi\\ &r_{12}=r_{12\max},r_{12\min} \end{aligned}\right\} \tag{24}$$

或

2．对同号的r_1,r_2，还有

$$\left.\begin{aligned} \varphi^*&=\arccos\left(\frac{r_1^2+r_2^2}{2r_1r_2}-\sqrt{\left(\frac{r_1^2+r_2^2}{2r_1r_2}\right)^2-1}\right)\\ r_{12}^*&=\frac{r_1^2+r_2^2}{2r_1r_2}-\sqrt{\left(\frac{r_1^2+r_2^2}{2r_1r_2}\right)^2-1} \end{aligned}\right\} \tag{25}$$

或

3．对异号的r_1,r_2，还有

$$\left.\begin{aligned} \varphi^{**}&=\arccos\left(\frac{r_1^2+r_2^2}{2r_1r_2}+\sqrt{\left(\frac{r_1^2+r_2^2}{2r_1r_2}\right)^2-1}\right)\\ r_{12}^{**}&=\frac{r_1^2+r_2^2}{2r_1r_2}+\sqrt{\left(\frac{r_1^2+r_2^2}{2r_1r_2}\right)^2-1} \end{aligned}\right\} \tag{26}$$

或

因此，当r_1,r_2同号时，K_2有三个可能极值点，它们是$\varphi=0,\pi,\varphi^*$或$r_{12}=r_{12\max},r_{12\min},r_{12}^*$；当$r_1,r_2$异号时，$K_2$也有三个可能极值点，它们是$\varphi=0,\pi,\varphi^{**}$或$r_{12}=r_{12\max},r_{12\min},r_{12}^{**}$。

分析表明，上述可能极值点均是K_2的极值点。对r_1,r_2的不同组合，这些点上K_2的极值性质(极大、极小)和取值由表1给出。由表1知：(1) K_2的极大值只出现在r_{12}的极值点上($r_{12\max}$或$r_{12\min}$上)；(2) r_{12}^*,r_{12}^{**}上K_2恒取极小值；(3) K_2的极大值均为1，极小值均为$\max(r_1^2,r_2^2)$。

图3 a—c给出了表1所列五种类型的一些K_2曲线。↑号指示了K_2极小值位置。

我们首先根据图3分析$r_{12}=0$(即因子间线性无关)时K_2的取值规律。发现，当且仅

表 1 K_2 极值点的位置、性质和取值

分类	r_1、r_2 特征		极大值		极小值	
	符号	数值	位置 (r_{12})	K_2	位置 (r_{12})	K_2
a	同号	$r_1、r_2 \neq 0$	$r_{12\min}^{\triangle}, r_{12\max}$	1	r_{12}^{*}	$\max(r_1^2, r_2^2)$
c	异号	$\lvert r_1\rvert \neq \lvert r_2\rvert$	$r_{12\max}^{\triangle}, r_{12\min}$	1	r_{12}^{**}	$\max(r_1^2, r_2^2)$
b	同号	$r_1、r_2 \neq 0$	$r_{12\min}^{\triangle}$	1	$r_{12}^{*}=r_{12\max}=1$	$\max(r_1^2, r_2^2)=r^2$
d	异号	$\lvert r_1\rvert=\lvert r_2\rvert=r$	$r_{12\max}^{\triangle}$	1	$r_{12}^{**}=r_{12\min}=-1$	$\max(r_1^2, r_2^2)=r^2$
e	$\max(\lvert r_1\rvert, \lvert r_2\rvert)=r$ $\min(\lvert r_1\rvert, \lvert r_2\rvert)=0$		$r_{12\max}, r_{12\min}$	1	$r_{12}^{*}=r_{12}^{**}=0$	$\max(r_1^2, r_2^2)=r^2$

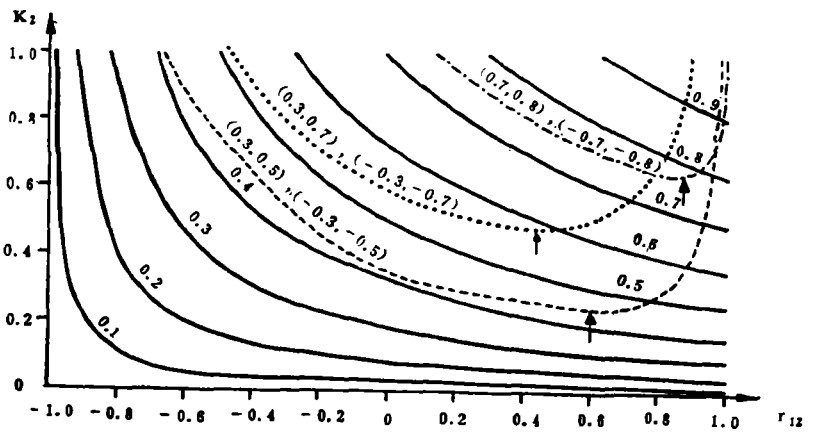


图 3a a(虚线)、b(实线)类 K_2 曲线

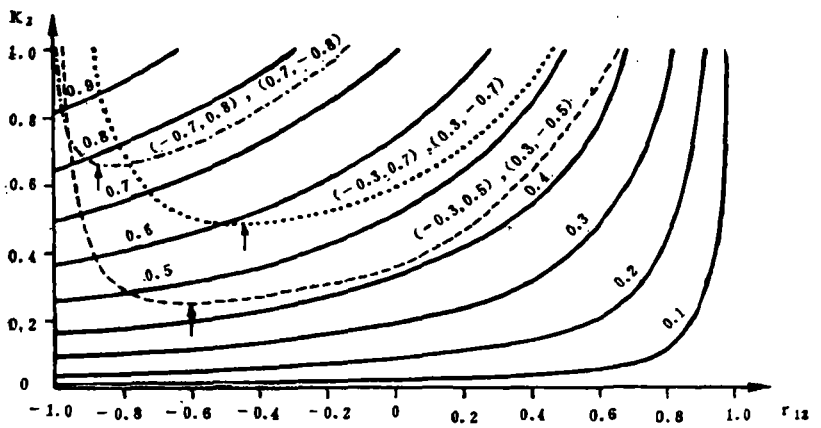


图 3b c(虚线)、d(实线)类 K_2 曲线

当 $r_1^2+r_2^2=1$ 时 K_2 才取得极大值 1(图 3a,b);否则,当 $r_1^2+r_2^2>1$ 时,由制约关系(10),不存在 $r_{12}=0$,因此,也就不存在 $K_2|_{r_{12}=0}$;当 $r_1^2+r_2^2<1$ 时,$K_2=r_1^2+r_2^2<1$,它不是 K_2 的极大值;特别地,对 e 类情形(图 3c),$K_2|_{r_{12}=0}$ 取 K_2 的极小值。因此,从拟合优度曲线的形状看,传统的选因子标准一般不能保证选到最优的一对因子。在样本容量较大的情

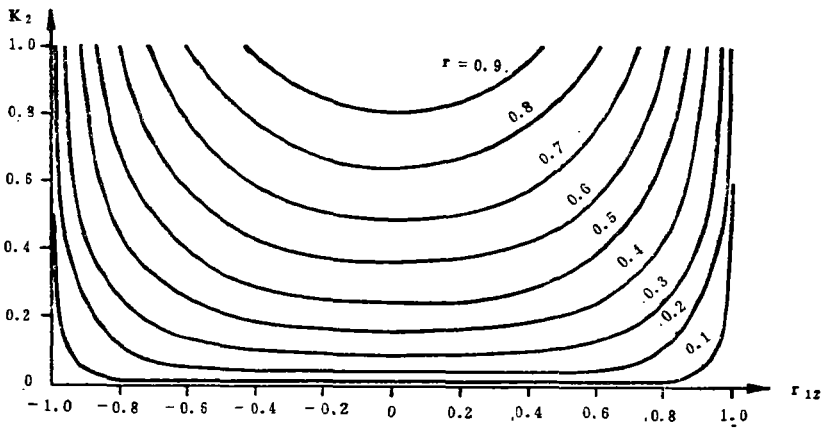


图 3c　e 类 K_2 曲线

况下，我们能够选到的因子，其与预报量间的相关系数一般不易满足 $r_1^2+r_2^2\geqslant 1$ 的条件[5]。此时，按传统的标准选因子建立起来的二元线性回归方程，其拟合优度只可能处于一个较低的水平。

假定我们所研究的气象过程可以看作平稳随机过程，则其统计特征将不随过程的推移而变，作为统计特征之一的拟合优度也就是回归方程的预报优度。因此，我们对于提高回归方程的拟合优度是极感兴趣的。于是，对二元线性回归方程，就存在一个挑选最优因子对(使 K_2 达到 1 的一对因子)的问题。最优因子对的位置已由表 1 极大值栏给出。

考察图 3 a，b，发现其上 a，c 类曲线(虚线)两端邻域内斜率绝对值有显著差异，a 类右端大于左端，c 类左端大于右端。这一差异随 $r_1^2+r_2^2$ 和 $|r_1^2-r_2^2|$ 之缩小而增大，极端地，当 $r_1^2=r_2^2$ 时，虚线退化为相应图上的实线，其右(左)端之极大值退化为极小值。这种情况使我们有可能从 a，c 类的两个 K_2 极大值(它给出 2 个最优因子对)中区别出一个相对稳定的最优因子对，称它为稳定最优因子对。而对于 b，d 类，保留下来的一个 K_2 极大值其右(左)邻域内 K_2 曲线的斜率绝对值较小，也称它为稳定最优因子对。稳定最优因子对的 r_{12} 取值已用符号“△”标注于表 1 上。在实际工作中，我们处理的大量问题是属于 $r_1^2+r_2^2$ 和 $|r_1^2-r_2^2|$ 均较小的问题，并且，r_{12} 的变动是不可避免的。因此，从最优因子对中区分出稳定最优因子对的工作是有重要实际意义的。

由稳定最优因子对性质出发，引出一个选择二元线性回归方程最优因子的标准：

(1) 因子与预报量间相关系数绝对值要尽可能大；

(2) 当因子与预报量间相关系数符号相同时，因子与因子间相关系数的代数值要尽可能小；当因子与预报量间相关系数符号相异时，因子与因子间相关系数的代数值要尽可能大。

这个标准的第一点与传统的选因子标准相同；第二点在 $r_1^2+r_2^2\geqslant 1$ 时(小样本时会迁到)与传统的选因子标准相同，在 $r_1^2+r_2^2<1$ 时(大样本时普遍存在)与传统标准相异。

这个标准的合理性可以从下面两方面得到论证：首先，它是将因子对作为整体看待的，因此包含了它们之间联系对预报量的作用；而传统的选因子标准则立足于单因子的作

用，因此排除了因子间相互联系对预报量的作用。已有的天气学知识支持新的选因子标准。例如，上游关键区出现适当的温压场配置(例如，斜压扰动)时，预报区域未来出现某种天气的机会要比其上单独出现上述配置中同类温度场或同类气压场时预报区域未来出现该种天气的机会多得多。这种情况表明，物理模型中温压场是作为一个整体与预报区域未来天气相联系的，尽管关键区中温度场和气压场间存在密切的联系，我们也没有理由将其中之一排除在统计模型之外。其次，新的选因子标准排斥将同一时刻(时段)邻近两点的同一要素选作因子对，因为，要素的空间代表性往往使 r_1，r_2 同号而使 r_{12} 取较大的正值；同理，新的选因子标准排斥将同一地点(地区)相近时刻(时段)的同一要素选作因子对。因此，新的选因子标准能够保证所选因子对包含较多的预报信息。

四、制约关系对多元线性回归方程拟合优度的影响

在讨论因子间的相关联系对多元线性回归方程拟合优度 K_p 的影响前，必须指出：(1) 因子数 p 远小于样本容量 n；(2) 因子间相关系数按(10)式受制于因子与预报量间的相关系数。

从(22)式出发的讨论将导致一系列实际计算，无法一般地揭示制约关系对 K_p 的影响，文献[2,4]中包括大量这方面的例子。为此，设 Y 与所有因子间的相关系数均为 r，所有预报因子间的相关系数均为 r'。此时，制约关系(10)取形式

$$r'=r^2+(1-r^2)\cos\varphi \tag{27}$$

而(19)、(20)给出的行列式为

$$D(r,r';p)=\begin{vmatrix} 1 & r & r & \cdots & r \\ r & 1 & r' & \cdots & r' \\ r & r' & 1 & \cdots & r' \\ & & \cdots\cdots & & \\ r & r' & r' & \cdots & 1 \end{vmatrix} \tag{28}$$

$$D_0(r';p)=\begin{vmatrix} 1 & r' & \cdots & r' \\ r' & 1 & \cdots & r' \\ & & \cdots & \\ r' & r' & \cdots & 1 \end{vmatrix} \tag{29}$$

计算行列式 D，D_0 的递推公式为

$$\left.\begin{aligned} D(r,r';p)&=(1-r^2)(1-s_1)^{p-1}\prod_{i=1}^{p-1}(1+s_i) \\ s_{i+1}&=\frac{s_i}{1+s_i},\ s_1=\frac{r'-r^2}{1-r^2} \end{aligned}\right\} \tag{30}$$

$$\left.\begin{aligned} D_0(r';p)&=(1-t_1)^{p-1}\prod_{i=1}^{p-1}(1+t_i)\\ t_{i+1}&=\frac{t_i}{1+t_i},\ t_1=r'\end{aligned}\right\} \tag{31}$$

利用(27)、(30)、(31)式将(22)式化作

$$K(r,\varphi;p)=1-(1-r^2)\prod_{i=1}^{p-1}\frac{1-i\cos\varphi}{1-(i-1)\cos\varphi}\Big/\frac{1+i[r^2-(1-r^2)\cos\varphi]}{1+(i-1)[r^2-(1-r^2)\cos\varphi]} \tag{32}$$

此式给出了 K_p 与 r,φ,p 的关系，或由(27)式，它给出了 K_p 与 r,r',p 的关系。不失一般性[3]，讨论 $r\in(0,1)$、$\varphi(r')$ 如表 2 所给值时 K_p 随 p 变化的特征。表 2 中 1，3，5 点对应 r' 取 r 所允许的极小值，平均值、极大值；2 点对应因子间完全不相关的情形，只对 $r\leqslant\frac{\sqrt{2}}{2}$ 成立；4 点对应因子间相关程度与因子和预报量间相关程度相等的情况。

表 2　　r' 的五个特性点

点　序	φ	r'	因子间相关特性
1	π	$2r^2-1$	极 小 值
2	$\arccos\left(-\frac{r^2}{1-r^2}\right)$	0	完全不相关
3	$\frac{\pi}{2}$	r^2	平 均 值
4	$\arccos\frac{r}{1+r^2}$	r	与 r' 相当
5	0	1	极 大 值

将表 2 中五点的 φ 值代入(32)式并化简，得

1. $K_p=1,p=2$ (33)

2. $K_p=pr^2,p\leqslant\frac{1}{r^2}$ (34)

3. $K_p=\frac{pr^2}{pr^2+(1-r^2)}$ (35)

4. $K_p=\frac{pr^2}{pr+(1-r)}$ (36)

5. $K_p=r^2$ (37)

对 $r=0.5$，用(33)—(36)式计算了表 2 所列五种情形下的 K_p 曲线。

图 4 表明，对于确定的 r 和 p，多元线性回归方程的拟合效果将因因子间相关系数的增大而变差。其中，第 3 点的 φ 或 r' 具有临界值的意义，当 $\varphi>\frac{\pi}{2}$ 或 $r'<r^2$ 时，理论上存

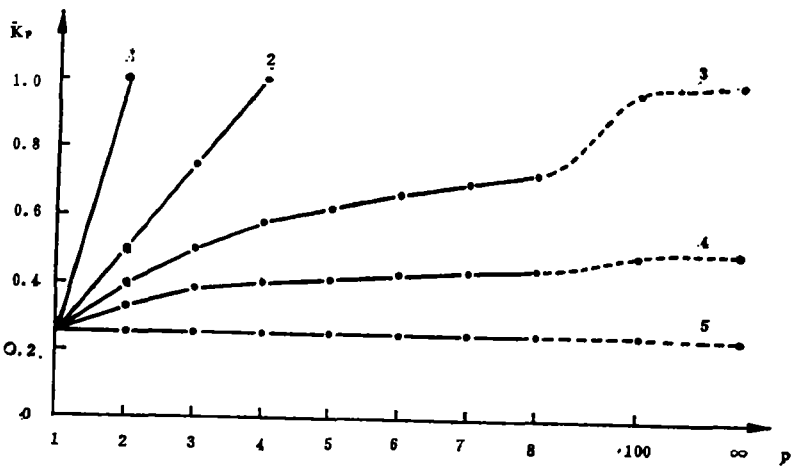


图 4 不同 r' 对应的 K_r 曲线($r=0.5$)

在以有限个因子达到拟合率 1 的可能性,反之,则因子的增加将不可能使 K_r 达 1。例如,对情况 4,即使无限增加因子,其 K_r 值也不可能超过 r(本例为 0.5)。

实际工作中经常迁到的样本对于取得高相关系数的因子而言似乎是太大了[5],而对于增多回归因子来说又似乎是太小了[4]。例如,对于一个容量为 62 的样本,得到一个 $r>0.330$ 的因子约需查找 100 对关系;而对这样的样本,因子数取到 6 已是相当大了。此时,若严格按照传统的选因子标准去选取预报因子,将不可避免地去普查数以千计的关系,而所得多元线性回归方程的拟合优度也不过略高于 0.5。因此,我们平时所见到的许多复回归方程,其较高的拟合优度已经包含了因子间相互作用的贡献。

五、结 语

探讨提高多元线性回归方程拟合优度的途径是近年来许多学科对多元分析提出的一个重要研究课题[4]。在实际工作中,传统的选因子的标准已不全为人遵守,一些回归方案中并不包括限制因子间相关的条件。本文的工作指出,传统的选因子标准有可修改之处;但是,由于理论上存在不稳定最优因子对,它对于一些仅根据复相关系数大小取舍因子的回归方案是一种隐患,因此,这类方案都是不完好的。在天气统计学领域内,如下两方面的问题也许更具有实质性。其一是把天气演变过程作为平稳随机过程看待是否符合客观实际,已有一些文献[6,7],讨论了这个问题;其二是线性联系是否是天气演变过程前后联系的普遍形式,文献[8]对这个问题进行过讨论。本文是在过程平稳和线性相关联系存在的假定下讨论多元线性回归方程回归因子间的合理关系的。

参 考 文 献

[1] 王宗皓、李麦村等编著,天气预报中的概率统计方法,科学出版社,65—75,88—101,1974.
[2] 黄嘉佑编,气象统计预报试用教材,北京大学地球物理系气象专业印,21—100,1976.

[3] 施能，论变量的相关性在组合因子时的作用，南京气象学院学报，1980 年第 1 期，28—33。
[4] M·肯德尔著，多元分析，中国科学院计算中心概率统计组译，科学出版社，1—15，108—138，1983。
[5] 徐国昌，亚洲 500 毫巴月平均高度的均方差分布及其在长期预报中的应用，气象学报，第 35 卷，第 3 期，333—337，1965。
[6] 林学椿，统计天气预报中相关系数的不稳定性问题，大气科学，第 2 卷，第 1 期，55—63，1978。
[7] 王绍武，美国第七届气候诊断年会，气象科技，1984 年第 4 期，54—61。
[8] 黄伯英，对目前概率预报中有关因子挑选和组合的几点讨论，气象科技资料，1977 年，7 期，8—10。

ON THE RELATIONSHIP AMONG THE LINEAR CORRELATION COEFFICIENTS OF VARIABLES AND ITS INFLUENCE ON GOODNESS OF FIT IN MULTIVARIATE LINEAR REGRESSION EQUATION

Wang Panxing

(Nanjing Institute of Meteorology)

Abstract

A quantitative study of the relationship among the correlation coefficients in the case of 3 variables is presented. Based on these relations the influence of the factor correlations on goodness of fit in multivariate regression equation is examined. At last a criterion is suggested for choosing the factors and reasonableness of the criterion is discussed.