

约束选择指数的通径分析及决策分析

王丽波¹, 孙洪罡², 刘璐¹, 郭满才¹, 袁志发¹

(1 西北农林科技大学 生命科学学院, 陕西 杨凌 712100; 2 中国农业大学 理学院, 北京 100083)

[摘要] 建立了约束选择指数的通径分析化模型, 分析其实质为约束选择指数的相关遗传力分解模型。进而运用通径分析的决策分析方法对约束选择指数进行决策分析, 定义各性状对聚合遗传值的决策系数, 各性状的决策系数代表其对聚合遗传值的综合决定作用。通过各性状的决策系数可以确定主选性状、限制性状和辅助性状, 使育种者在育种时, 可直接选择个体的表型就能达到使复合育种值进展最大的目的。

[关键词] 约束选择指数; 通径分析; 相关遗传力; 决策系数

[中图分类号] S813; O212

[文献标识码] A

[文章编号] 1671-9387(2006)03-0033-04

Hazel 1943 年提出了综合选择指数, Kempthorne 等 1959 年提出了约束选择指数, Tallis 1962 年提出了最宜选择指数, Resek 等 1969 年提出了理想选择指数^[1]。在国内, 袁志发等^[2]提出了选择指数与相关遗传进展的分解原理, 实现了各选择性状的相关遗传进展分解, 但仍未解决选择指数的最优路径问题。周静芋等^[3]提出了约束选择指数的通径分析化方法, 但其复合育种值并未标准化, 其实质仍为约束选择指数的相关遗传进展分解原理。2001 年袁志发等^[4]又提出了通径分析中的决策指标——决策系数, 用其可以将各自变量对响应变量的综合作用进行排序, 以确定主要决策变量和限制性变量。本文建立了约束选择指数的通径分析化模型, 分析得出其实质为约束选择指数的相关遗传力分解原理。进而应用通径分析的决策分析法对约束选择指数进行了决策分析, 给出了确定主选性状、限制性状和辅助性状的方法。该分析使得育种者可以直接选择表型而不用计算每个个体的选择指数值, 即可找到使聚合遗传值进展最大的最优路径。

1 约束选择指数的通径分析化模型

选择性状为 $x_1, x_2, \dots, x_m$, 记为 $X = (x_1, x_2, \dots, x_m)^T \sim N_m(\mu, \Sigma_p)$; X 的育种值为 $g = (g_1, g_2, \dots, g_m)^T \sim N_m(\mu, \Sigma_g)$; X 的环境离差值为 $e = (e_1, e_2, \dots, e_m)^T \sim N_m(0, \Sigma_e)$, 则 $X = g + e$ 。设 g 与 e 独立, 则有 $\Sigma_p = \Sigma_g + \Sigma_e$ 。其中, $\Sigma_p(\sigma_{p_{ij}})_{m \times m}$, $\Sigma_g(\sigma_{g_{ij}})_{m \times m}$ 和 $\Sigma_e(\sigma_{e_{ij}})_{m \times m}$, 分别为 X 的表型协方差阵、遗传协方差阵

和环境协方差阵。

设选择指数和复合育种值分别为:

$$I = a_1x_1 + a_2x_2 + \dots + a_mx_m = a^T X, \quad (1)$$

$$H = \omega_1g_1 + \omega_2g_2 + \dots + \omega_mg_m = \omega^T g, \quad (2)$$

式中, $a = (a_1, a_2, \dots, a_m)^T$, 为待估计系数; $\omega = (\omega_1, \omega_2, \dots, \omega_m)^T$, 为已知的经济权重列向量。 I 的表型方差、遗传方差分别为:

$$\begin{aligned} \sigma_{I_p}^2 &= a^T \Sigma_p a; \\ \sigma_{I_g}^2 &= a^T \Sigma_g a; \end{aligned} \quad (3)$$

H 的方差为:

$$\sigma_{H}^2 = \omega^T \Sigma_g \omega; \quad (4)$$

I 与 H 的协方差为

$$\text{Cov}(I, H) = a^T \Sigma_g \omega = \omega^T \Sigma_g a; \quad (5)$$

I 与 H 的相关系数为:

$$r_{IH} = \frac{a^T \Sigma_g \omega}{\sqrt{(a^T \Sigma_p a)(\omega^T \Sigma_g \omega)}}. \quad (6)$$

确定 a 的原则是在前 r 个性状的遗传进展为已知的条件下, 使相关系数 r_{IH} 达到最大, 故可转化为如下条件极值问题:

$$\begin{cases} a^T \Sigma_p a = 1, \\ a^T \Sigma_g C = l^T, \\ a^T \Sigma_g \omega = \max. \end{cases} \quad (7)$$

式中, $C = \begin{pmatrix} I_r \\ 0 \end{pmatrix}_{m \times r}$, I_r 为 r 阶单位阵, $l = (l_1, l_2, \dots, l_r)^T$ 。

应用拉格朗日乘数法, 令目标函数为

[收稿日期] 2005-07-26

[作者简介] 王丽波(1980—), 女, 黑龙江大庆人, 在读硕士, 主要从事农业应用数学研究。

[通讯作者] 袁志发(1938—), 男, 陕西渭南人, 教授, 博士生导师, 主要从事农业应用数学研究。

$$f = a^T \Sigma_k \omega - \frac{\alpha}{2} (a^T \Sigma_p a - 1) - (a^T \Sigma_k C - l^T) \lambda, \quad (8)$$

式中, α 和 λ 分别为拉格朗日乘数法的系数和列向量。令 $\frac{df}{da} = 0$, 得

$$\Sigma_k \omega - \alpha \Sigma_p a - \Sigma_k C \lambda = 0. \quad (9)$$

由于 α 为标量, a 为向量, 故 a 可由

$$a = \Sigma_p^{-1} \Sigma_k (\omega - C \lambda) \quad (10)$$

确定, 而不改变 a 各分量的比值。

又由式(7)中 $a^T \Sigma_k C = l^T$, 可得

$$C^T \Sigma_k a = l. \quad (11)$$

将式(10)代入式(11), 得

$$C^T \Sigma_k \Sigma_p^{-1} \Sigma_k (\omega - C \lambda) = l, \quad (12)$$

整理式(12)得

$$\lambda = (C^T \Sigma_k \Sigma_p^{-1} \Sigma_k C)^{-1} (C^T \Sigma_k \Sigma_p^{-1} \Sigma_k \omega - l). \quad (13)$$

令 $\omega' = \omega - C \lambda$, 则式(10)可转化为

$$\Sigma_p a = \Sigma_k \omega'. \quad (14)$$

式中, $\omega' = (\omega'_1, \omega'_2, \dots, \omega'_m)^T$, 为由于有了约束而改变了的经济权重, 即经济权重是 ω 的约束选择指数, 可看作是经济权重变为 ω' ($\omega' = \omega - C \lambda$) 的综合选择指数; λ ($\lambda = (C^T \Sigma_k \Sigma_p^{-1} \Sigma_k C)^{-1} (C^T \Sigma_k \Sigma_p^{-1} \Sigma_k \omega - l)$) 的生物学意义是由于有了约束而产生的被约束性状的经济权重的改变量。

设经济权重改变后的复合育种值为

$$H' = \omega'_1 g_1 + \omega'_2 g_2 + \dots + \omega'_m g_m = \omega'^T g. \quad (15)$$

由于 g 与 e 独立, 故

$$\text{Cov}(x_i, H') = \text{Cov}(g_i + e_i, \sum_{j=1}^m \omega'_j g_j) = \sum_{j=1}^m \omega'_j \sigma_{g_i},$$

所以式(14)可写为

$$\Sigma_p a = \text{Cov}(x, H'), \quad (16)$$

式中, $\text{Cov}(x, H') = (\text{Cov}(x_1, H'), \text{Cov}(x_2, H'), \dots, \text{Cov}(x_m, H'))^T$ 。

对式(16)表示方程组的各方程两边分别除以 $\sigma_{p1} \sigma_{H'}, \sigma_{p2} \sigma_{H'}, \dots, \sigma_{pm} \sigma_{H'}$, 并令 $b_j = \sigma_{pj} a_j / \sigma_{H'}$, 其中, σ_{pj} 为第 j 个性状的标准差; $\sigma_{H'}$ 为经济权重改变后育种值的标准差, 则式(16)可转化为

$$R_p b = R_{iH'}, \quad (17)$$

式中, $R_p = (r_{pij})_{m \times m}$, 为 $x_1, x_2, \dots, x_m$ 的表型相关阵; $b = (b_1, b_2, \dots, b_m)^T$ 为途径系数; $R_{iH'} = (r_{i1H'}, r_{i2H'}, \dots, r_{imH'})^T$, 为 X 与 H' 的相关阵。

式(17)即为约束选择指数的通径分析化模型, 即选择性状 X 与经济权重 ω' 的复合育种值 H' 的相关系数分解模型。对于式(7)而言, 若 $r = 0$, 则为综

合选择指数; 若 $r \neq 0, l = 0$, 则为约束选择指数; 若 $r \neq 0, l \neq 0$, 则为最宜选择指数。该模型将性状 x_i 与复合育种值 H' 的相关系数剖分为直接路 $x_i \rightarrow H'$ 对 H' 的直接贡献 b_i , 和 $m-1$ 条表型相关路 $x_i \xleftrightarrow{r_{pij}} x_j \rightarrow H' (j \neq i)$ 对 H' 的间接贡献 $r_{pij} b_j$ 之和。

该模型不同于以往的通径分析化模型, 以往的通径分析化模型两边均为表型相关或遗传相关, 笔者认为其应用在选择指数法上是不适合的, 不符合选择的本质, 因为运用选择指数法选择的本质, 是通过选择表型让复合育种值获得最大的进展。文中给出的综合选择指数的通径分析化模型是混合相关系数模型, 左边是表型相关, 右边是表型与遗传相关, 事实上由于 g 与 e 独立, 故有

$$r_{iH'} = \frac{\text{Cov}(x_i, H')}{\sigma_{p_i} \cdot \sigma_{H'}} = \frac{\text{Cov}(g_i + e_i, H')}{\sigma_{p_i} \cdot \sigma_{H'}} = \frac{\text{Cov}_g(x_i, H')}{\sigma_{p_i} \cdot \sigma_{H'}}. \quad (18)$$

又由相关遗传力的定义可知, 性状 x_i 与复合育种值 H' 的相关遗传力为

$$h_{iH'} = \frac{\text{Cov}_g(x_i, H')}{\sigma_{p_i} \cdot \sigma_{H'}}. \quad (19)$$

比较式(18)和式(19)可知, $r_{iH'} = h_{iH'}$ 。即性状 x_i 与复合育种值 H' 的表型相关系数等于 x_i 与 H' 的相关遗传力, 故式(17)可转化为

$$R_p b = h_{iH'}. \quad (20)$$

式中, $h_{iH'} = (h_{i1H'}, h_{i2H'}, \dots, h_{imH'})^T$, 为性状 X 与复合育种值 H' 的相关遗传力矩阵。

式(20)说明约束选择指数的通径分析化模型的实质是, 选择性状 X 与改变经济权重后的复合育种值 H' 的相关遗传力的分解模型。

在通径分析式(17)中, $x_1, x_2, \dots, x_m$ 对 H' 的决定系数为

$$R^2 = \sum_{i=1}^m b_i^2 + 2 \sum_{i < j} b_i r_{pij} b_j = \sum_{i=1}^m R_i^2 + \sum_{i < j} R_{ij}. \quad (21)$$

式中, $R_i^2 = b_i^2$, 为 x_i 对 H' 的直接决定系数; $R_{ij} = 2b_i r_{pij} b_j$, 为 x_i 对 H' 的相关决定系数(亦称间接决定系数), 显然 $R_{ij} = R_{ji}$ 。通径分析给出了各个性状与复合育种值 H' 间详细的路径信息, 如果能确定每个性状对复合育种值的综合作用, 那么就可以确定主选性状、限制性状和辅助性状。进而可以解决相关性状 $x_1, x_2, \dots, x_m$ 与复合育种值 H' 的最优路径问题, 约束选择指数的决策分析能解决上面提出的问题。

2 约束选择指数的决策分析

2.1 决策系数及其分析

所谓性状 x_i 对复合育种值 H' 的决策系数 $R_{(i)}$, 表示 x_i 对 H' 的综合决定能力, 即

$$R_{(i)} = R_i^2 + \sum_{j \neq i} R_{ij} = b_i^2 + \sum_{j \neq i} 2b_i r_{pij} b_j = 2h_{x_i H'} b_i - b_i^2. \quad (22)$$

式中, $R_{(i)}$ 表示性状 x_i 对复合育种值 H' 的综合决定作用, 既包括 x_i 通过直接路 $x_i \rightarrow H'$ 对 H' 的直接作用, 也包括 x_i 通过 $m-1$ 条表型相关路 $x_i \xleftrightarrow{r_{pij}} x_j \rightarrow H' (j \neq i)$ 对 H' 的间接作用。将 $R_{(1)}, R_{(2)}, \dots, R_{(m)}$ 由大到小排序, 设决策系数中最大者为 $R_{(i)}$ 。若 $R_{(i)} > 0$, 说明 x_i 对 H' 的综合促进作用最大, 故 x_i 为主选性状, 其他决策系数大于 0 的性状可作为辅助性状; 如果决策系数 $R_{(k)} < 0$, 说明 x_k 对 H' 起限制作用, 故 x_k 为限制性状, 其中最小决策系数对应的是主要限制性状。根据各性状的决策系数可以确定主选性状、辅助性状和限制性状, 这样就找到了使复合育种值进展最大的最优选择路径, 从而使得育种者可以直接选择个体的表型。

2.2 相关参数计算公式

(1) 令 $\sigma_p = \text{diag}\{\sqrt{\sigma_{p11}}, \sqrt{\sigma_{p22}}, \dots, \sqrt{\sigma_{pmm}}\}$, 则 $R_p = \sigma_p^{-1} \Sigma_p \sigma_p^{-1} p$ 。

$$(2) C = \begin{pmatrix} I_r \\ 0 \end{pmatrix}_{m \times r}, \lambda = (C^T \Sigma_k \Sigma_p^{-1} \Sigma_k C)^{-1}$$

$$(C^T \Sigma_k \Sigma_p^{-1} \Sigma_k \omega - l), \omega' = \omega - C\lambda, a = \Sigma_p^{-1} \Sigma_k \omega'.$$

(3) X 与 H' 的相关系数: $r_{x_i H'} = \sigma_p^{-1} \Sigma_k \omega' / \sigma_{H'}$ 。

(4) 通径系数: $b = R_p^{-1} r_{x_i H'}$ 。

(5) I 对 H' 的决定系数:

$$R_{IH'}^2 = \frac{a^T \Sigma_p a}{\omega'^T \Sigma_k \omega'} = \frac{a^T \Sigma_k \omega'}{\omega'^T \Sigma_k \omega'}.$$

(6) H' 的相关遗传进展:

$$CGS_{H'(I)} = k \sigma_{H'} = k \sqrt{a^T \Sigma_p a},$$

其中, k 为选择强度; $\sigma_{H'}$ 为 I 的表型标准差。

(7) X 的相关遗传进展:

$$CGS_{x(i)} = k \Sigma_k a / \sqrt{a^T \Sigma_p a}.$$

(8) x_i 对 H' 的经济贡献率:

$$\eta_i = \omega_i CGS_{x(i)} / CGS_{H'(I)}.$$

3 应用举例及其分析

在鸡的选择中, Kempthorne 等用 4 个性状制定了鸡的约束选择指数^[1]。4 个性状分别是: 成年体重(x_1)、初产日龄(x_2)、卵重(x_3)和至 72 周产卵数的 1/3(x_4), 表型协方差阵和遗传协方差阵分别为

$$\Sigma_p = \begin{bmatrix} 34 & 0 & 6.775 & -2.1988 \\ 0 & 13 & 0 & -10.8771 \\ 6.775 & 0 & 21.6 & -1.7526 \\ -2.1988 & -10.8771 & -1.7526 & 56.88 \end{bmatrix},$$

$$\Sigma_k = \begin{bmatrix} 15.3 & 0 & 3.855 & 1.9783 \\ 0 & 5.2 & 0 & -3.8457 \\ 3.855 & 0 & 10.8 & 1.6626 \\ 1.9783 & -3.8457 & 1.6626 & 11.376 \end{bmatrix}$$

各性状的经济权重 $\omega = (-2.50, 0, 7.20, 10.80)^T$ 。

表 1 约束选择指数的通径分析及决策分析结果

Table 1 Path analysis and decision analysis of restricted selection index

项目 Item	通径系数 Coefficient of path	相关系数 Correlation coefficient	决策系数 Decision coefficient
无约束 Nonrestraint	$b_1 = -0.0517$	$r_{x_1 H'} = 0.6034$	$R_{(1)} = -0.0653$
	$b_2 = -0.1140$	$r_{x_2 H'} = -0.2579$	$R_{(2)} = 0.0548$
	$b_3 = 0.4455$	$r_{x_3 H'} = 0.4147$	$R_{(3)} = 0.1711$
	$b_4 = 0.3596$	$r_{x_4 H'} = 0.2688$	$R_{(4)} = 0.0640$
对 x_1 和 x_2 约束 Restrict x_1 and x_2	$b_1 = -0.1882$	$r_{x_1 H'} = -0.0886$	$R_{(1)} = -0.0021$
	$b_2 = 0.1348$	$r_{x_2 H'} = -0.0177$	$R_{(2)} = -0.0229$
	$b_3 = 0.4747$	$r_{x_3 H'} = 0.4086$	$R_{(3)} = 0.1626$
	$b_4 = 0.3813$	$r_{x_4 H'} = 0.3130$	$R_{(4)} = 0.0933$
对 x_3 约束 Restrict x_3	$b_1 = -0.1010$	$r_{x_1 H'} = -0.0818$	$R_{(1)} = 0.0063$
	$b_2 = 0.1305$	$r_{x_2 H'} = -0.0361$	$R_{(2)} = -0.0265$
	$b_3 = -0.0108$	$r_{x_3 H'} = -0.0809$	$R_{(3)} = 0.0016$
	$b_4 = 0.4166$	$r_{x_4 H'} = 0.3712$	$R_{(4)} = 0.1357$

由表 1 可知, 无约束时, x_3 的决策系数最大且为正, 说明卵重对聚合遗传值进展起主要促进作用, 故卵重为主选性状; x_1 的决策系数最小且为负, 说

明成年体重对聚合遗传值进展起主要限制作用, 故成年体重为主要限制性状; x_2 和 x_4 的决策系数亦为正, 说明初产日龄和至 72 周产卵数的 1/3 也对聚

合遗传值进展起促进作用,故这二者可作为辅助选择性状。对 x_1 和 x_2 约束时, x_3 的决策系数最大且为正,说明此时卵重为主选性状; x_2 的决策系数最小且为负,说明此时初产日龄为主要限制性状;此时至 72 周产卵数的 1/3 和成年体重分别为辅助选择性状和限制性状。对 x_3 约束时, x_4 的决策系数最大且为正,说明此时至 72 周产卵数的 1/3 为主选性状;

x_3 的决策系数约等于 0,说明卵重对聚合遗传值进展几乎不起作用;此时 x_2 的决策系数为负,说明初产日龄为限制性状。

由表 2 可知,综合选择指数和约束选择指数相比, $CGS_{x(t)}$ 和 R_{III}^2 均变小,即不宜约束过多性状。值得注意的是,若约束的是限制性状,二者减小的幅度小;若约束的是主选性状,二者减小的幅度大。

表 2 约束选择指数的相关参数值

Table 2 Value of correlated parameter

参数 Parameter	无约束 Nonrestraint	对 x_1 和 x_2 约束 Restrict x_1 and x_2	对 x_3 约束 Restrict x_3
λ^T		(2.096 8, -7.482 9)	(8.703 7)
ω^T	(-2.5, 0.7, 2, 10.8)	(-4.596 8, 7.482 9, 7.2, 10.8)	(-2.5, 0. -1.503 7, 10.8)
b^T	(-0.051 7, -0.114 1, 0.445 5, 0.359 6)	(-0.188 2, 0.134 8, 0.474 7, 0.381 3)	(-0.088 1, -0.179 5, -0.007 7, 0.345 4)
R_{III}^2	0.350 6	0.327 6	0.212 6
$CGS_{H(t)}$	26.450 4k	23.507 6k	16.720 2k
$CGS_{x(t)}$	(0.554 5, -0.587 4, 1.824 6, 1.361)k	(0.0, 1.856 8, 0.938 7)k	(-0.318 6, -0.943 4, 0.1.474 4)k
$\eta_i/\%$	-5.24, 0.49, 67.55, 57	0, 0.56, 87.43, 13	4.76, 0, 0.95, 24

4 结 论

本文建立了约束选择指数的通径分析化模型,分析其实质为性状与聚合遗传值的相关遗传力分解

模型。运用通径分析的决策分析方法,对模型进行分析,给出了确立主选性状、限制性状及辅助选择性状的方法。该方法使得育种者可以通过直接选择表型就能找到使聚合遗传值进展最大的最优路径。

[参考文献]

- [1] 吴仲贤,统计遗传学[M].北京:科学出版社,1979:333-352.
- [2] 袁志发,常智杰,刘光祖,选择指数与相关遗传进展的分解原理[J].西北农业大学学报,1988,16(4):31-34.
- [3] 周静芋,孙世铎,宋世德,等.约束选择指数的通径分析化模型[J].西北农业大学学报,1995,23(5):98-103.
- [4] 袁志发,周静芋,郭满才,等.决策系数——通径分析中的决策指标[J].西北农林科技大学学报:自然科学版,2001,29(5):131-133.

Path analysis and decision analysis of restricted selection index

WANG Li-bo¹, SUN Hong-gang², LIU Lu¹, GUO Man-cai¹, YUAN Zhi-fa¹

(¹ College of Life Sciences, Northwest A & F University, Yangling, Shaanxi, 712100, China;

² College of Science, China Agricultural University, Beijing 100083, China)

Abstract: Restricted selection index, restricted selection index was treated as comprehensive selection index with economic weight changed in order to analyze its essence. The path analysis model of restricted selection index was established, and the essence of the path analysis model of restricted selection index was correlated heritability analysis model between traits and aggregate genetic economic value. Furthermore, the decision analysis of restricted selection index was fulfilled by using decision analysis method of path analysis. Decision coefficient of every trait, representing comprehensive effect of it on aggregate genetic economic value was defined. And primary selection trait together with auxiliary and restricted traits was determined by decision coefficient of every trait. Thus, the best selection strategy to maximize the aggregate genetic gain is found, and it allows breeders to select phenotypes directly.

Key words: restricted selection index; path analysis; correlated heritability; decision coefficient