

表型方差最大主成分性状的决策分析^{*}

刘璐, 王丽波, 郭满才, 袁志发

(西北农林科技大学 生命科学学院, 陕西 杨凌 712100)

[摘要] 采用通径分析的决策分析方法, 对表型方差最大主成分性状建立了通径分析化模型, 进而完成了对其的决策分析。文中给出了判断主成分中主载性状、限制性状及中间性状的方法, 便于育种者根据个体的性状表现来选择, 以达到对 F_i 选择所期望得到的遗传进展。

[关键词] 表型; 组合性状; 主成分; 通径分析; 决策分析

[中图分类号] S11⁺ 4; O212.4

[文献标识码] A

[文章编号] 1671-9387(2005)10-0097-03

对于多个性状的选择, 刘垂开等^[1]曾提出表型方差最大主成分性状, 育种者运用这种方法确定的组合性状, 在计算主成分值后可进行选择。为了使育种者不计算主成分值而直接通过预选个体的性状进行选择, 本文用袁志发等^[2,3]提出的通径分析中的决策分析方法, 解决了表型方差最大主成分性状的最优路径问题, 给出了主成分性状中的主选性状、辅助性状和限制性状的确定方法, 从而可使育种者直接根据预选个体的性状表现进行选择, 进而通过对个体主成分性状的选择而达到育种目的。

1 表型方差最大主成分性状及其选择效果

设要选择的性状为 $X = (x_1, x_2, \dots, x_m)^T \sim N_m(\mu, \Sigma_p)$, X 的育种值向量和环境离差向量分别为 $g = (g_1, g_2, \dots, g_m)^T \sim N_m(\mu, \Sigma_g)$ 和 $e = (e_1, e_2, \dots, e_m)^T \sim N_m(0, \Sigma_e)$, 则 $X = g + e$ 。若 g 与 e 相互独立, 则有 $\Sigma_p = \Sigma_g + \Sigma_e$, 其中 $\Sigma_p, \Sigma_g, \Sigma_e$ 分别为 X 的表型、遗传型和环境协方差阵。

令 $\mu_i = (\mu_{1i}, \mu_{2i}, \dots, \mu_{mi})^T$, 则表型方差最大主成分性状为一组合性状:

$$F_i = u_{1i}x_1 + u_{2i}x_2 + \dots + u_{mi}x_m = u_i^T X \quad (1)$$

其中表型、遗传和环境方差分别为:

$$\sigma_{F_i}^2 = u_i^T \Sigma_p u_i, \sigma_{F_i, g}^2 = u_i^T \Sigma_g u_i, \sigma_{F_i, e}^2 = u_i^T \Sigma_e u_i \quad (2)$$

u_i 应满足:

$$\begin{cases} u_i^T u_i = 1 \\ u_i^T \Sigma_p u_i \text{ 最大} \end{cases} \quad (3)$$

用拉格朗日(Lagrange)乘数法解之, u_i 应满足:

$$(\Sigma_p - \lambda_i I_m) u_i = 0 \quad (4)$$

由于 Σ_p 实对称且正定, 故特征方程: $|\Sigma_p - \lambda I_m| = 0$ 的解有:

$$\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_m > 0 \quad (5)$$

λ_i 对应的特征向量为 $u_i = (u_{1i}, u_{2i}, \dots, u_{mi})^T$ 。令 $U = (u_1, u_2, \dots, u_m)^T$ 且知 U 为正交矩阵, 故有:

$$F = \begin{bmatrix} F_1 \\ F_2 \\ \vdots \\ F_m \end{bmatrix} = \begin{bmatrix} u_{11} & u_{21} & \dots & u_{m1} \\ u_{12} & u_{22} & \dots & u_{m2} \\ \vdots & \vdots & \ddots & \vdots \\ u_{1m} & u_{2m} & \dots & u_{mm} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_m \end{bmatrix} = U^T X \quad (6)$$

式中, $F_1, F_2, \dots, F_m$ 分别称为表型方差最大第1个, 第2个..., 第 m 个主成分性状, 其有如下性质:

λ_i 是 F_i 的表型方差, 即 $\sigma_{F_i}^2 = u_i^T \Sigma_p u_i = \lambda_i$;

各不同 F_i 间相互独立, 即 $\text{cov}(F_i, F_j) = 0$, $i \neq j$ 。

表型方差最大主成分性状 F_i 的遗传力和直接遗传进展分别为:

$$h_{F_i}^2 = \frac{u_i^T \Sigma_g u_i}{u_i^T \Sigma_p u_i}, GS_{F_i} = k \sigma_{F_i, p}^2 h_{F_i}^2 \quad (7)$$

由于 F_i 的直接遗传进展使各 x_j 的相对遗传进展为:

$$GGS_{j(F_i)} = \frac{\text{cov}(g_j, F_{ig})}{u_i^T \Sigma_g u_i} GS_{F_i} = \frac{u_{1i} \sigma_{g_1} + u_{2i} \sigma_{g_2} + \dots + u_{mi} \sigma_{g_{im}}}{u_i^T \Sigma_g u_i} GS_{F_i}$$

* [收稿日期] 2005-03-10

[作者简介] 刘璐(1976-), 女, 陕西宝鸡市人, 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。

$$\frac{k}{\sqrt{\lambda_i}} (u_{1i}\sigma_{g1} + u_{2i}\sigma_{g2} + \dots + u_{mi}\sigma_{gm})$$

$$\text{故 } (CGS_{1(F_i)}, CGS_{2(F_i)}, \dots, CGS_{m(F_i)}) = \frac{k}{\sqrt{\lambda_i}} \Sigma u_i \quad (8)$$

2 表型方差最大主成分性状的通径分析化模型

由式(4)知,表型方差最大主成分性状 F_i 应满足: $\Sigma u_i = \lambda u_i$, 即

$$\begin{bmatrix} \sigma_{p1}^2 & \sigma_{p12} & \dots & \sigma_{p1m} \\ \sigma_{p21} & \sigma_{p2}^2 & \dots & \sigma_{p2m} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{pm1} & \sigma_{pm2} & \dots & \sigma_{pm}^2 \end{bmatrix} \begin{bmatrix} u_{1i} \\ u_{2i} \\ \vdots \\ u_{mi} \end{bmatrix} = \begin{bmatrix} \lambda u_{1i} \\ \lambda u_{2i} \\ \vdots \\ \lambda u_{mi} \end{bmatrix} \quad (9)$$

对第 j 个方程两边同除以 $\sigma_{pj}\sqrt{\lambda}$, 并令 $u_{ji} = \sigma_{pj}u_{ji}/\sqrt{\lambda}$, 并由(6)式变形得 $X = UF$, 有 $x_j = u_{j1}F_1 + u_{j2}F_2 + \dots + u_{jm}F_m$, 且 x_j 与 F_i 的相关系数为:

$$r_{x_j F_i} = \frac{\text{cov}(F_i, u_{j1}F_1 + u_{j2}F_2 + \dots + u_{jm}F_m)}{\sigma_{pj}\sigma_{F_i}} = \frac{\sqrt{\lambda} u_{ji}}{\sigma_{pj}} \quad (10)$$

式中, $j = 1, 2, \dots, m$, 则式(9)变为:

$$\begin{cases} u_{1i} + r_{p12}u_{2i} + \dots + r_{p1m}u_{mi} = r_{x_1 F_i} \\ r_{p21}u_{1i} + u_{2i} + \dots + r_{p2m}u_{mi} = r_{x_2 F_i} \\ \vdots \\ r_{pm1}u_{1i} + r_{pm2}u_{2i} + \dots + u_{mi} = r_{x_m F_i} \end{cases} \quad (11)$$

$$\text{即} \begin{bmatrix} 1 & r_{p12} & r_{p13} & \dots & r_{p1m} \\ r_{p21} & 1 & r_{p23} & \dots & r_{p2m} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ r_{pm1} & r_{pm2} & r_{pm3} & \dots & 1 \end{bmatrix} \begin{bmatrix} u_{1i} \\ u_{2i} \\ \vdots \\ u_{mi} \end{bmatrix} = \begin{bmatrix} r_{x_1 F_i} \\ r_{x_2 F_i} \\ \vdots \\ r_{x_m F_i} \end{bmatrix} \text{ 或 } R_P U_i = R_{X F_i} \quad (12)$$

式中, R_P 为 X 的表型相关阵; $R_{X F_i}$ 为 X 与 F_i 的相关列向量。

式(11)或(12)即为表型方差最大主成分性状 F_i 的通径分析模型。其遗传育种意义为: 性状 x_j 与 F_i 的表型相关系数 $r_{x_j F_i}$ 可分解为 m 个部分: x_j 的直接贡献为 u_{ji} ; x_j 与 x_k ($k \neq j$)对 $r_{x_j F_i}$ 的贡献为 $r_{pj k}u_{ki}$, 其通径为 $x_j \leftarrow \frac{r_{pj k}}{x_k} \leftarrow \frac{u_{ki}}{F_i}$, $j = 1, 2, \dots, m$, $k \neq j$, 共

有 $m-1$ 项, 当 $m=3$ 时, 其通径图见图1。

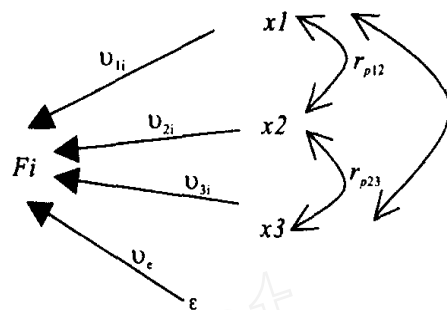


图1 X 对 F_i 的通径图

Fig 1 Path graph of X to F_i

3 表型方差最大主成分性状 F_i 的决策分析

F_i 的通径分析化模型给出了各性状对 $r_{x_j F_i}$ 的直接贡献和间接贡献, 但没有给出对 F_i 的综合贡献, 故无法评定 X 中的各性状对 F_i 的主次作用。袁志发等^[2,3]在研究通径分析时, 解决了这个问题。据此, 本研究解决了各性状 x_j 对表型方差最大主成分性状 F_i 的综合决定作用问题, 此种分析称为表型方差最大主成分性状的决策分析或最优路径分析。具体方法如下所述。

X 中各性状对 F_i 的决定系数可分解为:

$$R_{X F_i}^2 = \sum_{j=1}^m u_{ji} r_{x_j F_i} = \sum_{j=1}^m (\sigma_{pj} u_{ji} / \sqrt{\lambda}) \times (\sqrt{\lambda} u_{ji} / \sigma_{pj}) = \sum_{j=1}^m u_{ji}^2 = 1 \quad (13)$$

这说明 X 中各性状对 F_i 的决定是完全的。然而 $R_{X F_i}^2$ 又可分解为:

$$R_{X F_i}^2 = \sum_{j=1}^m u_{ji} r_{x_j F_i} = \sum_{j=1}^m u_{ji} (r_{p j 1} u_{1i} + r_{p j 2} u_{2i} + \dots + r_{p j m} u_{mi}) = u_{1i}^2 + u_{2i}^2 + \dots + u_{mi}^2 + 2 \sum_{j < k} u_{ji} r_{p j k} u_{ki} = 1 \quad (14)$$

式(14)表示 X 对 F_i 的总决定系数为1, 其可分解为 x_j 与 F_i 的直接决定系数 $R_j^2 = u_{ji}^2$ (m 个)和 $x_k \leftarrow \frac{r_{p k j}}{x_j} \leftarrow x_l$ 的相关决定系数 $R_{kl} = R_{lk} = 2u_{ki} r_{p k l} u_{li}$ (共 C_m^2 项)。

由式(14)可知, x_j 与 F_i 的综合决定能力即决策系数为:

$$R_{i(j)} = R_j^2 + \sum_{k \neq j} R_{kj} = 2u_{ji} r_{x_j F_i} - u_{ji}^2 \quad (15)$$

对 $R_{i(j)}$ 从大到小排序, 就可看出 X 中各性状对 F_i 的综合决定作用的大小, 其中为正值且最大者为决定 F_i 选择效果的主选性状; 最小者(为负)为限制 F_i 选择效果的限制性性状; $R_{i(j)}$ 位次处于最大和最小中间

的性状可相继判断其对 F_i 选择效果的作用, 要么辅助主选性状, 要么作适当的限制。有了决策系数描述各性状对 F_i 的综合决定能力后, 育种者就可按决策分析的结果对预选个体的性状表现直接选择, 而不需要计算这个个体的 F_i 值。

$$\Sigma_p = \begin{bmatrix} 40 & 094 & 9 & - & 14 & 359 & 4 & 1 & 110 & 2 & - & 20 & 977 & 5 & - & 4 & 902 & 1 & 12 & 608 & 5 & 3 & 911 & 4 \\ - & 14 & 359 & 4 & 48 & 417 & 3 & 0 & 707 & 3 & 12 & 419 & 1 & 7 & 923 & 5 & 15 & 192 & 0 & - & 6 & 736 & 0 \\ 1 & 110 & 2 & 0 & 707 & 3 & 10 & 819 & 9 & 1 & 691 & 1 & 0 & 927 & 1 & - & 15 & 188 & 3 & - & 0 & 975 & 4 \\ - & 20 & 977 & 5 & 12 & 419 & 1 & 1 & 691 & 1 & 50 & 561 & 7 & 10 & 644 & 6 & 45 & 701 & 5 & 8 & 662 & 2 \\ - & 4 & 902 & 1 & 7 & 923 & 5 & 0 & 927 & 1 & 10 & 644 & 6 & 8 & 914 & 0 & 0 & 930 & 3 & 1 & 891 & 7 \\ 12 & 608 & 5 & 15 & 192 & 0 & - & 16 & 960 & 9 & 45 & 701 & 5 & 0 & 930 & 3 & 113 & 870 & 0 & 1 & 746 & 4 \\ 3 & 911 & 4 & - & 6 & 736 & 0 & - & 0 & 975 & 4 & 8 & 662 & 2 & 1 & 891 & 7 & 1 & 746 & 4 & 2 & 743 & 9 \end{bmatrix} \begin{matrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \\ x_6 \\ x_7 \end{matrix}$$

经分析, 第一表型方差最大主成分性状为:

$\lambda_1 = 143\ 591\ 7$, F_1 对 7 个性状的方差贡献率为

$$F_1 = -0.0191x_1 + 0.2015x_2 - 0.0912x_3 + 0.4599x_4 + 0.0546x_5 + 0.8575x_6 + 0.0301x_7$$

51.17%, 据式(11)和(12), F_1 的通径分析化模型为:

$$\begin{bmatrix} 1 & 0.000 & 0 & - & 0.325 & 9 & 0.053 & 3 & - & 0.465 & 9 & - & 0.259 & 3 & 0.186 & 6 & 0.372 & 9 \\ - & 0.325 & 9 & 1 & 0.000 & 0 & 0.030 & 9 & 0.251 & 0 & 0.381 & 4 & 0.204 & 6 & - & 0.584 & 4 \\ 0.053 & 3 & 0.030 & 9 & 1 & 0.000 & 0 & 0.072 & 3 & 0.094 & 4 & - & 0.432 & 7 & - & 0.179 & 0 \\ - & 0.465 & 9 & 0.251 & 0 & 0.072 & 3 & 1 & 0.000 & 0 & 0.501 & 4 & 0.602 & 3 & 0.735 & 4 \\ - & 0.259 & 3 & 0.381 & 4 & 0.094 & 4 & 0.501 & 4 & 1 & 0.000 & 0 & 0.029 & 2 & 0.382 & 5 \\ 0.186 & 6 & 0.204 & 6 & - & 0.483 & 2 & 0.602 & 3 & 0.029 & 2 & 1 & 0.000 & 0 & 0.098 & 8 \\ 0.372 & 9 & - & 0.584 & 4 & - & 0.179 & 0 & 0.735 & 4 & 0.382 & 5 & 0.098 & 8 & 1 & 0.000 & 0 \end{bmatrix} \begin{bmatrix} u_1 \\ u_2 \\ u_3 \\ u_4 \\ u_5 \\ u_6 \\ u_7 \end{bmatrix} = \begin{bmatrix} - & 0.036 & 1 \\ 0 & 347 & 0 \\ - & 0.332 & 2 \\ 0 & 775 & 0 \\ 0 & 219 & 1 \\ 0 & 962 & 9 \\ 0 & 217 & 7 \end{bmatrix}$$

由 $U_{j1} = \sigma_{pjuj1} / \sqrt{\lambda_1}$ 求得:

$$(U_{11}, U_{21}, \dots, U_{71}) = (-0.0016, 0.1170, -0.0250, 0.2729, 0.0136, 0.7636, 0.0042)$$

据公式(15)可计算出 X 的各性状对 F_1 的决策系数: $R_{1(1)} = 0.00013$, $R_{1(2)} = 0.0658$, $R_{1(3)} = 0.01599$, $R_{1(4)} = 0.34852$, $R_{1(5)} = 0.00577$, $R_{1(6)} = 0.88746$, $R_{1(7)} = 0.00181$ 。

从以上结果可以看出, 各个性状对 F_1 的决策系数排序为: $R_{1(6)} > R_{1(4)} > R_{1(2)} > R_{1(3)} > R_{1(5)} > R_{1(7)} > R_{1(1)}$ 且未出现负值, 这说明 F_1 主要承载的是 x_6 和 x_4 , 即旗叶面积和株高因子, 其他 5 个性状对其没有构成限制。穗粒数(x_1)和有效穗数(x_5)二者与千粒重(x_2)的选择方向相反, 因而从表型上选择旗叶面积大、植株高、有效穗数少的植株即可达到 F_1 的选择进展。

4 应用实例

下面以小麦的 7 个主要经济性状的表型方差最大主成分性状的决策分析为例说明之。数据来自参考文献[4]。数据的表型协方差阵为:

5 讨论

本文对表型方差最大主成分性状结合通径分析的决策分析进行了深入研究。用各性状 x_i 对 F_i 的决策系数 $R_{i(j)}$ 判断性状 x_i 对 F_i 的综合决定作用的大小; 用 $R_{i(j)}$ 排序结果分析 F_i 中的主载性状、限制性状和可利用性状, 从而给出 F_i 的具体形象; 用 F_i 的具体形象可直接从个体的性状表现中选择个体, 达到 F_i 的选择进展。其次, 因各 F_i 间相互独立, 可对总方差贡献率大于 90% 的前几个主成分性状独立进行决策分析, 找出其各自的具体形象, 然后综合一个总的性状形象, 以表示所研究性状 $x_1, x_2, \dots, x_m$ 中的主要变异来源, 便于决定群体利用方向。另外, 表型方差最大的主成分性状的遗传力或遗传方差并不一定是最大, 因而本方法可继续研究下去。

[参考文献]

- [1] 刘垂珏, 刘来福. 多个数量性状遗传分析的数据结构[J]. 遗传, 1985, 7(4): 12-14
- [2] 袁志发, 周静芋. 多元统计分析[M]. 北京: 科学出版社, 2002: 128-133
- [3] 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
- [4] 李邦发. 小麦主要经济性状遗传参数的研究[J]. 遗传, 1989, 11(1): 4-7

(下转第104页)

Study on a rapeseed huller based on two-time striking

GUO Gui-sheng, LU Xin-min, DANG Ge-rong, GUO Kang-quan

(College of Mechanism and Electronic Engineering, Northwest A & F University, Yangling, Shaanxi 712100, China)

Abstract: On the basis of analyzing the hulling methods and principles for general agricultural crop materials, combined with physical and mechanical properties of rapeseed, a hulling method of two-time striking was developed. Hulling equipment was made according to the two-time striking principle, and the structure and principle of the equipment were introduced. Based on the equipment, test results showed that: hulling rate was more than 75% and powder rate was less than 5%. With an increase of rapeseed moisture content, hulling rate and powder rate were increased. The fixed position on delivering axis and height of the striking board had the best value.

Key words: two-time striking; rapeseed; huller

(上接第99页)

Abstract ID: 1671-9387(2005)10-0097-EA

The decision-making analysis of the biggest variance of phenotype traits with principal component

LIU Lu, WANG Li-bo, GUO Man-cai, YUAN Zhi-fa

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

Abstract: By using decision-making analysis of path analysis, this paper built a path analysis model for the biggest variance of phenotype traits with principal component and completed decision-making analysis and showed a way to judge the main weight-bearing trait, confinement trait and intermedial trait from principal component. So breeding operator can choose the traits from individual expression to achieve the expectation of genetic process by selection.

Key words: phenotype; combination character; principal component; path analysis; decision-making analysis