

网络出版时间:2013-01-25 17:27

网络出版地址: <http://www.cnki.net/kcms/detail/61.1390.S.20130125.1727.023.html>

决策系数的检验及在育种分析中的应用

解小莉,袁志发

(西北农林科技大学 理学院,陕西 杨凌 712100)

【摘要】 【目的】提出了检验决策系数的统计方法,为研究者分析不同自变量对因变量的综合作用提供了量化指标。【方法】采用推导随机变量函数均值和方差的通用方法—— Δ 技术,求解决策系数的均值和方差,提出决策系数的统计检验方法。【结果】建立了决策系数检验的 t 统计量,给出了检验的具体步骤。根据小麦产量与千粒质量、株穗数、穗粒数和穗粒质量 4 个影响因素的例子,得出了提高小麦产量必须显著地提高株穗数和限制穗粒数的结论。【结论】提出的决策系数统计检验方法,可以为分析不同自变量对因变量的综合作用提供量化指标。

【关键词】 决策系数; Δ 技术;统计检验

【中图分类号】 O212

【文献标志码】 A

【文章编号】 1671-9387(2013)03-0111-04

Statistical test of decision coefficient and its application in breeding

XIE Xiao-li, YUAN Zhi-fa

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

Abstract: 【Objective】 The study aimed to propose a statistical test method for examination of decision coefficient, which provided quantitative indexes for analyzing the comprehensive impact of independent variables on dependent variables. 【Method】 Δ method was used to compute the mean value and variance of the function of decision coefficient. 【Result】 The t statistic and test steps of the decision coefficient were established. Wheat yields were taken as an example to illustrate the statistical test of the decision coefficient. Results demonstrated that the panicle number per hill should be increased while the grain number per ear should be limited in order to increase the production of wheat. 【Conclusion】 The statistical test method for the decision coefficient was suggested and had the potential to be used to quantitate the impact of independent variables on dependent variables.

Key words: decision coefficient; Δ method; statistical test

通径分析是由数量遗传学家 Wright^[1] 于 1921 年提出的,现已广泛应用于自然、经济和社会等各个学科。通径分析是研究多个自变量 $x_1, x_2, \dots, x_p$ 与响应变量 y 之间多元线性回归关系的一种重要的统计方法。基于标准化的多元线性回归方程的正则方程组为:

$$\begin{cases} b_1^* + r_{12}b_2^* + \dots + r_{1p}b_p^* = r_{1y}, \\ r_{21}b_1^* + b_2^* + \dots + r_{2p}b_p^* = r_{2y}, \\ \vdots \\ r_{p1}b_1^* + r_{p2}b_2^* + \dots + b_p^* = r_{py}. \end{cases} \quad (1)$$

式中: b_j^* 是 x_j 的偏回归系数; r_{jk} 为 x_j 与 x_k 的相关系数; r_{jy} 为 x_j 与 y 的相关系数。

在通径分析中,分别定义了各自变量 x_j 对响应

【收稿日期】 2012-05-07

【基金项目】 08 回国人员科研启动基金项目(01140503)

【作者简介】 解小莉(1974—),女,陕西澄城人,讲师,博士,主要从事应用数学研究。E-mail: xiemary@nwsuaf.edu.cn

【通信作者】 袁志发(1938—),男,陕西渭南人,教授,博士生导师,主要从事应用数学研究。E-mail: zhifayuan@nwsuaf.edu.cn

变量 y 的直接决定系数 R_j^2 、间接决定系数 R_{jk}^2 ，以及自变量 $x_1, x_2, \dots, x_p$ 对响应变量 y 的决定系数 R^2 ，其具体公式为：

$$R_j^2 = b_j^{*2}; \quad (2)$$

$$R_{jk}^2 = 2b_j^* r_{jk} b_k^*; \quad (3)$$

$$R^2 = \sum_{j=1}^p R_j^2 + \sum_{j < k}^{p-1} R_{jk}^2 = \sum_{j=1}^p b_j^* r_{jy}. \quad (4)$$

通过多个自变量与响应变量的相关系数剖分，往往不容易确定哪个变量对响应变量起着决定性作用或主要限制性作用，针对这一问题，袁志发等^[2]于 2001 年提出了决策系数的概念，将自变量 x_j 对响应变量 y 的决策系数 $R_{(j)}$ 定义如下：

$$R_{(j)} = b_j^{*2} + 2 \sum_{j \neq k} b_j^* r_{jk} b_k^* = R_j^2 + \sum_{j \neq k} R_{jk}^2. \quad (5)$$

决策系数 $R_{(j)}$ 反映了自变量 x_j 通过 $x_1, x_2, \dots, x_p$ 的相关网对响应变量 y 的综合决定能力，不仅包含了直接决定系数 R_j^2 ，而且包含了与 x_j 有关的间接决定系数 $\sum_{j \neq k} R_{jk}^2 = 2 \sum_{j \neq k} b_j^* r_{jk} b_k^*$ 。

自从决策系数指标提出以后，大多数使用通径分析方法的文献都采用了该指标来进行决策分析^[3-13]。秦豪荣等^[3]根据决策系数分析得到：管围和膘厚是长白猪以瘦肉率为主要目标的典范选择指

$$\begin{cases} \frac{y_i - \bar{y}}{\sigma_y} = \beta_1^* \left(\frac{x_{i1} - \bar{x}_1}{\sigma_1} \right) + \beta_2^* \left(\frac{x_{i2} - \bar{x}_2}{\sigma_2} \right) + \dots + \beta_p^* \left(\frac{x_{ip} - \bar{x}_p}{\sigma_p} \right) + \frac{\epsilon_i}{\sigma_y}, \\ \beta_j^* = \beta_j \frac{\sigma_j}{\sigma_y}. \end{cases} \quad (7)$$

标准化多元线性回归方程的最小二乘估计式为

$$\begin{aligned} \frac{\hat{y} - \bar{y}}{S_y} &= b_1^* \left(\frac{x_1 - \bar{x}_1}{S_1} \right) + b_2^* \left(\frac{x_2 - \bar{x}_2}{S_2} \right) + \\ &\dots + b_p^* \left(\frac{x_p - \bar{x}_p}{S_p} \right). \end{aligned} \quad (8)$$

式中： $\bar{y}$ 、 S_y 为变量 y 的均值和标准差， $\bar{x}_i$ 、 S_i 为变量 x_i 的均值和标准差， $b_j^* = \frac{S_j}{S_y} b_j$ 为标准化偏回归系数， $j=1, 2, \dots, p$ 。

设 $\mathbf{b}^* = (b_1^*, b_2^*, \dots, b_p^*)^T$ ，则 $\mathbf{b}^*$ 的正则方程组及在 $\mathbf{R}_{xx}$ 可逆时的解为：

$$\mathbf{R}_{xx} \mathbf{b}^* = \mathbf{R}_{xy}, \mathbf{b}^* = \mathbf{R}_{xx}^{-1} \mathbf{R}_{xy}. \quad (9)$$

式中： $\mathbf{R}_{xx}$ 为 $x_1, x_2, \dots, x_p$ 的相关阵， $\mathbf{R}_{xy}$ 为 $x_1, x_2, \dots, x_p$ 与 y 的相关阵， $\mathbf{R}_{xx}^{-1} = (\mathbf{C}_{kt}^*)_{p \times p}$ ， $\mathbf{R}_{xx} \mathbf{b}^* = \mathbf{R}_{xy}$ 的具体形式为(1)式。

由标准化回归方程可以得到回归平方和为：

$$U^* = (b_1^*, b_2^*, \dots, b_p^*) \mathbf{R}_{xy} = \sum_{i=1}^p b_i^* r_{iy} = R^2. \quad (10)$$

剩余平方和及 σ^2 的估计分别为：

$$Q_e^* = 1 - U^*, \hat{\sigma}^2 = \frac{Q_e^*}{f_e} = \frac{1 - R^2}{f_e}. \quad (11)$$

数的主要决策变量。韩勇等^[4]利用决策系数分析得到：最外轮舌瓣长为标准切花菊花径大小的主选性状，最外轮舌瓣宽为标准切花菊花径大小的主要限制性状。决策系数的提出，解决了研究者正确地进行综合决策，以及确定响应变量的决策变量和限制性变量等问题。然而决策系数的检验至今尚未见报道。为此，本文采用推导随机变量函数均值和方差的通用方法—— Δ 技术，求解决策系数的均值和方差，提出决策系数的检验方法。

1 决策系数的统计检验

1.1 标准化的多元线性回归方程

对于观察点 $(x_{i1}, x_{i2}, \dots, x_{ip}, y_i)$ ($i=1, 2, \dots, n$)，自变量 $x_1, x_2, \dots, x_p$ 与响应变量 y 的一般多元线性回归方程为：

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_p x_{ip} + \epsilon_i. \quad (6)$$

其中 ϵ_i 相互独立且均服从 $N(0, \sigma^2)$ 。

对各变量进行标准化，设自变量 $x_1, x_2, \dots, x_p$ 与响应变量 y 的均值和方差分别为 $\mu_1, \mu_2, \dots, \mu_p, \mu_y$ 及 $\sigma_1, \sigma_2, \dots, \sigma_p, \sigma_y$ ，则标准化多元线性回归方程为：

式中： $f_e = n - p - 1$ ，为误差自由度。

1.2 Δ 技术^[14]

Δ 技术是推导随机变量函数的均值、方差和协方差的一种通用方法。如果随机变量 x 的均值为 μ ，方差为 σ^2 ，在计算 $g(x)$ 的均值和方差时，若 $g(x)$ 不是 x 的线性函数，则 $E(g(x)) \neq g(E(x))$ 。 $g(x)$ 在均值 μ 点的泰勒展开式为：

$$\begin{aligned} g &= g(x) = g(\mu) + \frac{dg(\mu)}{dx} (x - \mu) + \\ &\frac{d^2 g(\mu)}{2! dx^2} (x - \mu)^2 + \dots. \end{aligned} \quad (12)$$

如果对(12)等式两边取期望值，且忽略等式右边的第 3 项及更高次项，则得：

$$E(g) \approx g(\mu) + \frac{1}{2} \frac{d^2 g(\mu)}{dx^2} \sigma^2. \quad (13)$$

同样， g 的方差近似为：

$$V(g) \approx E(g - E(g))^2 = \left[\frac{dg(\mu)}{dx} \right]^2 \sigma^2. \quad (14)$$

1.3 决策系数的统计检验

ϵ_i 相互独立且均服从 $N(0, \sigma^2)$ ，则 $b_j^* \sim N(\beta_j^*,$

$C_{jj}^* \sigma^2)^{[15]}$,其中 C_{jj}^* 为 $\mathbf{R}_{xx}^{-1}$ 的主对角线元素, $R_{(j)}$ 是 b_j^* 的非线性函数,用 Δ 技术计算 $R_{(j)}$ 的期望值和方差分别为:

$$\begin{cases} E(R_{(j)})=2\beta_j^* r_{jy}-\beta_j^{*2}-C_{jj}^* \sigma^2, \\ V(R_{(j)})=4(r_{jy}-\beta_j^*)^2 C_{jj}^* \sigma^2. \end{cases} \quad (15)$$

由(11)式可以得到 σ^2 的估计,因此,对于无效假设 $H_{0j};E(R_{(j)})=0$,可近似用下列 t 检验:

$$t_j=\frac{|R_{(j)}|}{2|r_{jy}-b_j^*|\sqrt{C_{jj}^*\frac{1-R^2}{f_e}}}\sim t(f_e),j=1,2,\cdots,p. \quad (16)$$

其中, f_e 为误差自由度。

2 决策系数检验在育种分析中的应用

袁志发^[16]于 1981 年对小麦产量 y 与其构成因素 x_1 (千粒质量)、 x_2 (株穗数)、 x_3 (穗粒数)和 x_4 (穗粒质量)进行了通径分析。试验为随机区组设计,参试品种 10 个,重复 3 次。本文在通径分析的基础上,对决策系数进行了统计检验。试验数据估计的遗传相关系数见表 1。

表 1 小麦产量与其 4 个构成因素的遗传相关系数

Table 1 Genetic correlation coefficient of wheat yield and its four factors

指标 Index	千粒质量 x_1 Kernel weight x_1	株穗数 x_2 Panicle number per hill x_2	穗粒数 x_3 Grain number per ear x_3	穗粒质量 x_4 Grain weight per ear x_4	产量 y Yield y
千粒质量 x_1 Kernel weight x_1	1	0.274	-0.706	0.525	0.050
株穗数 x_2 Panicle number per hill x_2		1	-0.300	0.474	0.477
穗粒数 x_3 Grain number per ear x_3			1	-0.665	0.256
穗粒质量 x_4 Grain weight per ear x_4				1	0.440

通径分析的正则方程组为:

$$\begin{cases} b_1^*+0.274b_2^*-0.706b_3^*+0.525b_4^*=0.050, \\ 0.274b_1^*+b_2^*-0.300b_3^*+0.474b_4^*=0.477, \\ -0.706b_1^*-0.300b_2^*+b_3^*-0.665b_4^*=0.256, \\ 0.525b_1^*+0.474b_2^*-0.665b_3^*+b_4^*=0.440. \end{cases} \quad (17)$$

其解为:

$$\begin{aligned} b_1^*&=0.321, b_2^*=0.313, \\ b_3^*&=1.180, b_4^*=0.908. \end{aligned} \quad (18)$$

相关阵 $\mathbf{R}_{xx}$ 的逆矩阵为:

$$\begin{bmatrix} 1 & r_{12} & r_{13} & r_{14} \\ r_{21} & 1 & r_{23} & r_{24} \\ r_{31} & r_{32} & 1 & r_{34} \\ r_{41} & r_{42} & r_{43} & 1 \end{bmatrix}^{-1} = \begin{bmatrix} 2.0224 & -0.0910 & 1.2964 & -0.1565 \\ -0.0910 & 1.2946 & -0.0935 & -0.6280 \\ 1.2964 & 0.0935 & 0.6249 & 1.1093 \\ -0.1565 & -0.6280 & 1.1093 & 2.1175 \end{bmatrix}.$$

(19)

由式(10)和(11)得:

$$R^2=\sum_{i=1}^4 b_i^* r_{iy}=0.867, \hat{\sigma}^2=\frac{1-R^2}{18}=0.0074.$$

统计检验知: b_2^*, b_3^*, b_4^* 极显著, b_1^* 显著^[15]。

根据(5)式可以计算得到每个自变量对响应变量的决策系数 $R_{(j)}$, 结果见表 2。

表 2 小麦产量的通径分析和决策分析

Table 2 Path analysis and decision analysis of wheat yield

通径分析 Path analysis	直接作用 b_j^* Direct effect b_j^*	间接作用 $r_{jk}b_k^*$ Indirect effect $r_{jk}b_k^*$	总作用 Total effect	决策系数 $R_{(j)}$ Decision coefficient $R_{(j)}$
x_1 对 y x_1 to y	0.321	$x_1 \leftrightarrow x_2 \leftrightarrow y$ 0.086 $x_1 \leftrightarrow x_3 \leftrightarrow y$ -0.833 $x_1 \leftrightarrow x_4 \leftrightarrow y$ 0.477	0.050	-0.071
x_2 对 y x_2 to y		$x_2 \leftrightarrow x_1 \leftrightarrow y$ 0.088 $x_2 \leftrightarrow x_3 \leftrightarrow y$ -0.354 $x_2 \leftrightarrow x_4 \leftrightarrow y$ 0.430		
x_3 对 y x_3 to y		$x_3 \leftrightarrow x_1 \leftrightarrow y$ -0.227 $x_3 \leftrightarrow x_2 \leftrightarrow y$ -0.094 $x_3 \leftrightarrow x_4 \leftrightarrow y$ -0.604		
x_4 对 y x_4 to y	0.908	$x_4 \leftrightarrow x_1 \leftrightarrow y$ 0.169 $x_4 \leftrightarrow x_2 \leftrightarrow y$ 0.148 $x_4 \leftrightarrow x_3 \leftrightarrow y$ -0.785	0.440	-0.026

由(16)式对决策系数进行 t 检验,各变量的决策系数对应的 t 统计量分别为: $t_1=1.071, t_2=6.231, t_3=3.068, t_4=0.222$,自由度为 18,查表得

到显著临界值为: $t_{0.05}(18)=2.101, t_{0.01}(18)=2.878$,因而, $R_{(2)}$ 和 $R_{(3)}$ 极显著, $R_{(1)}$ 和 $R_{(4)}$ 不显著。综合决策分析得到:要提高小麦产量,必须显著地提

高株穗数(x_2),显著地限制穗粒数(x_3)。

3 结 论

本文在 Δ 技术的基础上,解决了决策系数的均值和方差问题,进一步提出了检验决策系数的统计方法。决策系数为研究者分析不同自变量对因变量的综合作用提供了量化指标,决策系数的统计检验方法的提出,使这一指标更加完善。最后,以小麦产量为育种目标,利用通径分析中的决策系数可以得到:显著地提高小麦的株穗数和限制小麦的穗粒数,可以达到提高小麦产量的目的。

[参考文献]

- [1] Wright S. Correlation and causation [J]. J Agricultural Research, 1921, 20: 557-585.
- [2] 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.
Yuan Z F, Zhou J Y, Guo M C, et al. Decision coefficient: The decision index of path analysis [J]. Jour of Northwest Sci-Tech Univ of Agri and For: Nat Sci Ed, 2001, 29(5): 131-133. (in Chinese)
- [3] 秦豪荣,袁志发,吉俊玲,等. 长白猪典范选择指数的构建与通径分析研究 [J]. 生物数学学报, 2007, 22(4): 679-684.
Qin H R, Yuan Z F, Ji J L, et al. Canoical selection index and its path analysis in Landrace pigs [J]. Journal of Biomathematics, 2007, 22(4): 679-684. (in Chinese)
- [4] 韩 勇,姚建军,陈发棣,等. 标准切花菊花径与其他重要数学性状的相关与通径分析 [J]. 南京农业大学学报, 2012, 35(1): 33-37.
Han Y, Yao J J, Chen F D, et al. Correlation and path analysis of flower diameter with other quantitative characters in standard cut chrysanthemum [J]. Journal of Nanjing Agricultural University, 2012, 35(1): 33-37. (in Chinese)
- [5] 南 娟,汪有科,李晓彬,等. 基于通径分析的枣树发芽率影响因素研究 [J]. 北方园艺, 2010(20): 10-13.
Nan J, Wang Y K, Li X B, et al. Study on factor of jujube rate based on path analysis [J]. Northern Horticulture, 2010(20): 10-13. (in Chinese)
- [6] 李艳艳,丰 震,赵兰勇,等. 玫瑰切花产量性状遗传参数和选择效率的初步研究 [J]. 园艺学报, 2007, 34(4): 955-958.
Li Y Y, Feng Z, Zhao L Y, et al. A preliminary study on genetic parameters and selection efficiency of yield traits of fresh flora bud in *Rosa rugose* [J]. Acta Horticulturae Sinica, 2007, 34(4): 955-958. (in Chinese)
- [7] 夏 磊,赵志新,汤 玲,等. 重庆地区加拿利海枣夏季光合日变化与主要环境因子的关系 [J]. 北京林业大学学报, 2011, 33(4): 75-80.
Xia L, Zhao Z X, Tang L, et al. Diurnal changes in summer photosynthetic rate of *Phoenix canariensis* in relation with environmental factors in Chongqing, Southern China [J]. Journal of Beijing Forestry University, 2011, 33(4): 75-80. (in Chinese)
- [8] 刘 璐,赵会仁,郭满才,等. 典范性状对的决策分析 [J]. 西北农林科技大学学报:自然科学版, 2009, 37(9): 182-186, 192.
Liu L, Zhao H R, Guo M C, et al. The decision-making analysis of the canonical pair of traits [J]. Jour of Northwest A&F Univ: Nat Sci Ed, 2009, 37(9): 182-186, 192. (in Chinese)
- [9] 张慧林,罗 军,刘小林,等. 西农萨能奶山羊泌乳均衡性与产奶量的相关分析 [J]. 西北农林科技大学学报:自然科学版, 2009, 37(9): 39-44.
Zhang H L, Luo J, Liu X L, et al. Correlation analysis between 300-day milk yield and proportionality in Xinong Saanen milk goat [J]. Jour of Northwest A&F Univ: Nat Sci Ed, 2009, 37(9): 39-44. (in Chinese)
- [10] 李云开,杨培岭,任树梅,等. 分形流道设计及几何参数对滴水水力性能的影响 [J]. 机械工程学报, 2007, 43(7): 109-114.
Li Y K, Yang P L, Ren S M, et al. Effects of fractal flow path designing and parameters on emitter hydraulic performance [J]. Chinese Journal of Mechanical Engineering, 2007, 43(7): 109-114. (in Chinese)
- [11] 王 芳,庞广昌,王景川. 螺旋藻 β -胡萝卜素代谢控制分析及其新方法的研究 [J]. 食品科学, 2011, 32(7): 237-243.
Wang F, Pang G C, Wang J C. Metabolic control analysis of β -carotene in *Spirulina* [J]. Food Science, 2011, 32(7): 237-243. (in Chinese)
- [12] 靳甜甜,傅伯杰,刘国华,等. 不同坡位沙棘光合日变化及其主要环境因子 [J]. 生态学报, 2011, 31(7): 1783-1793.
Jin T T, Fu B J, Liu G H, et al. Diurnal changes of photosynthetic characteristics of *Hippophae rhamnoid* and the relevant environment factors at different slope locations [J]. Acta Ecologica Sinica, 2011, 31(7): 1783-1793. (in Chinese)
- [13] 陈 鹏,叶 辉. 云南六库桔小实蝇成虫种群数量变动及其影响因素因子分析 [J]. 昆虫学报, 2007, 50(1): 38-45.
Chen P, Ye H. Population dynamics of *Bactrocera dorsalis* (Diptera: Tephritidae) in Liuku, Yunnan with an analysis of the influencing factors [J]. Acta Entomologica Sinica, 2007, 50(1): 38-45. (in Chinese)
- [14] 杨子恒. 计算分子进化 [M]. 钟 杨,张文娟,梅 旖,等译. 上海:复旦大学出版社, 2008.
Yang Z H. Computational molecular evolution [M]. Zhong Y, Zhang W J, Mi Q, et al translation. Shanghai: Fudan University Press, 2008. (in Chinese)
- [15] 袁志发,周静芋. 试验设计与分析 [M]. 北京:高等教育出版社, 2000.
Yuan Z F, Zhou J Y. Design and analysis of experiments [M]. Beijing: Higher Education Press, 2000. (in Chinese)
- [16] 袁志发. 通径分析方法简介 [J]. 麦类作物, 1981(3): 42-46, 48.
Yuan Z F. The introduction of path analysis [J]. Acta Tritical Crops, 1981(3): 42-46, 48. (in Chinese)