

决策系数——通径分析中的决策指标^{*}

袁志发, 周静芋, 郭满才, 雷雪芹, 解小莉

(西北农林科技大学 生命科学学院, 陕西 杨陵 712100)

[摘要] 提出了通径分析中的决策指标——决策系数, 用它可以把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。

[关键词] 通径分析; 决策系数; 决策指标排序

[中图分类号] S813 [文献标识码] A

[文章编号] 1000-2782(2001)05-131-03

通径分析是研究多个自变量 $x_1, x_2, \dots, x_p$ 与响应变量间线性关系的一种重要统计方法, 基本原理是基于标准化多元线性回归方程的正则方程组:

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

其中, b_i 是 x_i 的偏回归系数; r_{ij} 为 x_i 与 x_j 的相关系数; r_{iy} 为 x_i 与 y 的相关系数, $i, j = 1, 2, \dots, p$ 。

在方程组(1)中, 把 x_i 对 y 的总作用 r_{iy} 剖分为两部分: b_i 是 x_i 对 y 的直接作用, 它是 x_i 对 y ($x_i \rightarrow y$) 的直接通径系数; 其他 $p-1$ 项中, $r_{ij}b_j$ 是 x_i 通过 x_j 的相关路对 y ($x_i \leftarrow x_j \rightarrow y$) 的间接作用 ($i \neq j$), 或称为 x_i 通过 x_j 对 y 的间接通径系数, 即 $r_{iy} = b_i + \sum_{j \neq i} b_j r_{ij}$ 。

在通径分析中, 分别称

$$R_i^2 = b_i^2, R_{ij}^2 = 2b_i r_{ij} b_j \quad (2)$$

为 x_i 对 y 的直接决定系数和 x_i 与 x_j 通过相关路对 y 的间接决定系数。 $x_1, x_2, \dots, x_p$ 对 y 的决定系数为:

$$R^2 = \sum_{i=1}^p R_i^2 + \sum_{i=1}^{p-1} \sum_{j=i+1}^p R_{ij}^2 = \sum_{i=1}^p b_i r_{iy} \quad (3)$$

通径分析不但可以检验出各 x_i 对 y 的作用是否显著, 而且可以检验出 $x_1, x_2, \dots, x_p$ 对 y 的决定作用 (R^2) 是否显著, 最重要的是可以看出 x_i 是如何直接影响或如何通过其他自变量间接影响 y 的。然而, 通过多个自变量与 y 的相关系数的剖分, 往往很难明确看出那个变量对 y 起着主要决定性作用或主

要限制性作用, 特别对于专业不太熟悉的研究者更感困难。为此, 本研究提出各变量的决策系数, 以期更好的运用通径分析方法进行综合决策。

1 决策系数

1.1 定义

称

$$R_{(i)}^2 = b_i^2 + 2 \sum_{j \neq i} b_i r_{ij} b_j = R_i^2 + \sum_{j \neq i} R_{ij}^2 \quad (4)$$

为 x_i 对 y 的决策系数。

1.2 决策系数的统计意义

$R_{(i)}^2$ 反映了 x_i 通过 $x_1, x_2, \dots, x_p$ 的相关网对 y 的综合决定作用。它不仅包含了 x_i 对 y 的直接决定作用 R_i^2 , 还包含了与 x_i 有关的间接决定系数 $\sum_{j \neq i} R_{ij}^2 = 2 \sum_{j \neq i} b_i r_{ij} b_j$ 。间接决定系数 R_{ij}^2 既包含了 x_i 通过 x_j 对 y 的决定作用, 亦包含了 x_j 通过 x_i 对 y 的决定作用。具体分析有以下3点:

由(1)和(3)式知, $b_i r_{iy}$ 为 x_i 对 R^2 的贡献, 但

$$b_i r_{iy} = b_i^2 + \sum_{j \neq i} b_i r_{ij} b_j = R_{(i)}^2 - \sum_{j \neq i} b_i r_{ij} b_j \quad (5)$$

表明 $R_{(i)}^2$ 中还包含了除 x_i 之外的其他变量 x_j 通过 x_i 对 y 的间接作用。只有当 $r_{ij} = 0$ 时 $R_{(i)}^2 = R_i^2$ 。

由式(5)得:

$$R_{(i)}^2 = 2b_i r_{iy} - b_i^2 \quad (6)$$

式(6)表明 $R_{(i)}^2$ 未必为正值。若 $R_{(i)}^2 > 0$, 表明 x_i 对 y 起增进作用; 当 $R_{(i)}^2 < 0$, 表明 x_i 对 y 起限制性作用。

由式(3)和(4)有:

* [收稿日期] 2001-03-28

[作者简介] 袁志发(1938-), 男, 陕西渭南市人。教授, 博士生导师。主要从事农业应用数字的教学和研究。

$$\sum_{i=1}^p R_{(i)}^2 = R^2 + 2 \sum_{j=2}^{p-1} b_i r_{ij} b_j \quad (7)$$

表明决策系数之和一般并不等于 R^2 , 只有当所有 $r_{ij} = 0$ 时, 它们才相等。

由上述分析知, $R_{(i)}^2$ 和 $b_i r_{iy}$, R_i^2 及 R^2 是不同的, 它反映了 x_i 通过 $x_1, x_2, \dots, x_p$ 的相关网对 y 的综合作用。

利用 $R_{(i)}^2$ 值可以把各变量对 y 的综合作用由大到小排序, 排序最大的变量为主要决策变量, 但未必它的直接决定作用大; 排序最小的变量, 若其决策系

数为负, 则为主要限制性变量, 但未必它的直接决定作用小。

2 应用举例

袁志发^[5]于 1981 年对小麦产量 y 与其构成因素 x_1 (百粒重)、 x_2 (每株穗数)、 x_3 (每穗粒数) 和 x_4 (每穗粒重) 进行了通径分析。通过统计检验除 b_1 显著外, 其余标准化偏回归系数均极显著, $R^2 = 0.867$ 。它们的遗传相关系数见表 1。

表 1 小麦产量及其结构性状的遗传相关系数表

Table 1 Genetic correlation coefficient of wheat yield and structural characters

	x_1	x_2	x_3	x_4	y
x_1	1	0.274	-0.706	0.525	0.050
x_2		1	-0.300	0.474	0.477
x_3			1	-0.665	0.256
x_4				1	0.440

从相关上看, 仅 r_{2y} 显著, 但从生物学上看 x_1 和 x_3 亦是重要的, 只所以 r_{1y} 和 r_{2y} 小, 是由于 r_{13} , r_{23} 和 r_{34} 均是较强的负相关造成的, 其中详细情况可由通径分析结果(表 2)看出。

表 2 小麦产量通径分析表

Table 2 Path analysis of wheat yield

通径 Path	b_i	$r_{ij} b_j$	r_{iy}
x_1 对 y	0.321	x_2	0.086
		x_3	-0.833
		x_4	0.477
		x_1	0.088
x_2 对 y	0.313	x_3	-0.354
		x_4	0.430
		x_1	-0.227
		x_2	-0.094
x_3 对 y	1.180	x_4	-0.604
		x_1	0.169
		x_2	0.148
		x_3	-0.785
x_4 对 y	0.908		

由表 2 知, 直接作用 $b_3 > b_4 > b_1 > b_2$; 从与 y 的相关看, $r_{2y} > r_{4y} > r_{3y} > r_{1y}$; r_{1y} 只所以小(0.05), 是因为受 x_3 的限制($b_3 r_{13} = -0.833$)。在这种复杂的路径信息中, 选择什么样的路径对 y 最好呢? 决策系数可以解决这个问题。

由式(6)计算的各变量的决策系数为:

$$R_{(1)}^2 = 2 \times 0.321 \times 0.05 - 0.321^2 = -0.071$$

$$R_{(2)}^2 = 2 \times 0.313 \times 0.477 \times 0.313^2 = 0.200$$

$$R_{(3)}^2 = 2 \times 1.180 \times 0.256 - 1.180^2 = -0.790$$

$$R_{(4)}^2 = 2 \times 0.908 \times 0.440 - 0.908^2 = -0.026$$

按决策系数排序为: $R_{(2)}^2 > R_{(4)}^2 > R_{(1)}^2 > R_{(3)}^2$, 且 $R_{(3)}^2 < 0$ 。故 x_2 为主要决策变量(直接决定作用最

小), 其原因是它通过 x_1 和 x_4 对 y 的间接作用分别为 0.088 和 0.430, 协助 x_2 对 y 起增进作用, x_3 对它的限制最小(-0.354)。 x_3 为主要限制性变量(直接作用最大), 因为它通过其他变量的间接作用均为负。因此, 要提高产量, 必须提高每株穗数(x_2), 限制每穗粒数(x_3), 基本保持每穗粒重(x_4)或百粒重(x_1)。

3 讨论

本研究提出的决策系数为通径分析所提供的各种路径信息的综合决策理顺了思路。决策系数的排序结果反映了各变量对 y 的综合作用的大小, 它既

不同于偏回归系数的排序,也不同于各变量与 y 相关的排序。利用决策系数排序能很方便地确定各变量在决策中的作用,如确定主要决定性变量、主要限制性变量等。

[参考文献]

- [1] 西北农学院 概率基础与数理统计[M]. 北京: 中国农业出版社, 1988 237- 247.
- [2] 袁志发, 孟德顺 多元统计分析[M]. 陕西杨陵: 天则出版社, 1993 134- 143
- [3] 莫惠栋 农业试验统计[M]. 上海: 上海科学技术出版社, 1984 537- 545
- [4] 袁志发, 周静芋 试验设计与分析[M]. 北京: 高等教育出版社, 2000 187- 197.
- [5] 袁志发 通径分析[J]. 国外农学——麦类作物, 1981, (3): 30- 36

Decision coefficient—the decision index of path analysis

YUAN Zhi-fa, ZHOU Jing-yu, GUO Man-cai, LEI Xue-qin, XIE Xiao-li

(College of Life Science, Northwest Sci-Tech University of Agriculture and Forestry, Yangling, Shaanxi 712100, China)

Abstract: This paper advances a decision index of path analysis decision coefficient. By using it, sequence of the synthetic action of every independent variables to response variable can be decided which variable is the principal decision variable and which variable is the restricted variable can be identified.

Key words: path analysis; decision coefficient; the sequence of decision index

· 简 讯 ·

优质小麦新品种“小偃 926”通过审定

由西北农林科技大学农学院薛文江研究员主持选育的优质小麦新品种——“小偃 926”于 2001 年 8 月 29 日经陕西省农作物品种审定委员会审定通过。

该品种冬性, 分蘖力较强, 株型紧凑, 株高 80~ 85 cm, 抗倒伏; 早熟(比小偃 6 号早熟 2~ 3 d, 比陕 229 早熟 3~ 4 d), 叶功能期长, 灌浆速度快, 穗长方形, 结实性好, 成熟黄亮; 抗寒、耐旱、抗干热风、耐阴雨, 中抗叶枯病, 对条锈病抗扩展, 轻感赤霉和白粉病; 白粒、角质, 千粒重 40 g 左右。陕西省农产品质量监督检验站品质分析显示, 该品种蛋白质含量 17.34%, 湿面筋 38.88%, 吸水率 60.3%, 沉淀值 51.3 mL, 稳定时间 9 min, 磨粉品质好, 出粉率高, 属于优质面包小麦。

经过省区域试验和生产示范, 一般单产在 3 675 kg/hm² 以上, 最高可达 6 750 kg/hm²。适宜于关中塬灌区和旱肥地及同类型生产区域大面积种植。

(屈李纯 供稿)