● 学术交流丛书 ●

通 径 分 析

明道绪 编著

一九九〇年八月

通径分析

一九八〇年八月

通　径　分　析

明道绪　编著

内容提要

本书系统地介绍通径分析的原理和方法，包含编著者近年来在这方面的主要研究成果。全书共分四节，祥细介绍通径系数与决定系数等基本概念；通径系数与相关系数的关系；性状相关的通径分析方法以及通径分析的数学模型、显著性检验等内容。本书在从理论上阐述通径分析原理的同时，注意了与农业、畜牧业科学研究实践的联系，各节均有步骤完整、过程详细的实例。书末附有用电子计算器进行通径分析的手算程式与用 Basic 语言编制的电算程序。内容安排力求循序渐进，由浅入深，深入浅出，通俗易懂。本书可作为农业院校有关专业大、专学生和研究生的学习材料，也可供从事农、牧业科学研究和教学工作的同志参考。

前　言

作者从一九七八年底开始进行通径分析的学习、研究与教学。一九八二年四川农学院(现四川农业大学)教务处曾将作者与其合作者在通径分析研究上所撰写的几篇文章及一篇译文汇编成通径分析的原理与方法专辑印行(见四川农学院教务处编《科技资料》1982.2)。一九八五年作者应《农业科学导报》(现名《西南农业学报》)编辑部之约曾撰写《通径分析的原理与方法》一文在《农业科学导报》1986.1-4期上连载。但《科技资料》汇编的各篇文章因相对独立而连贯性较差;《农业科学导报》连载的文章因限于篇幅又失之过简。本书作者结合从事《通径分析》的教学经验，参阅了有关专著对上述两个资料进行了修改、充实编纂而成，重点在于介绍性状相关的通径分析及其显著性检验的原理与方法。

作者在从事《通径分析》的研究中，自始至终得到了我校高之仁教授、东北农学院盛志廉教授的大力支持和热情指导，在此向他们表示衷心的感谢!

本书在撰写过程中曾参阅了有关中外文献和专著，并引用了其中的一些资料，作者对这些文献和专著的作者们致以诚挚的谢意，对在本书出版过程中给予热情帮助的刘维庆、潘光堂同志表示感谢。

限于作者水平，书中难免有缺点和错误，敬请读者批评指正。

四川农业大学数量遗传研究室

明道绪

1990.3

绪　言

在动、植物遗传育种研究工作中，常常要研究多个数量性状间的关系，例如研究牛的体重与体长、胸围、腹围、体高的关系；研究禾谷类作物的单株生产力与每株穗数、每穗粒数、粒重、株高、穗长的关系等。从统计学的观点讲，就是要研究多个相关变量间的关系。相关变量的关系分为两类：因果关系与平行关系。在上述例子中，体重为果，体长、胸围、腹围、体高为因；单株生产力为果，每株穗数、每穗粒数、粒重、株高、穗长为因。在各原因变量间，它们是互为因果或受另一共同原因的作用而呈平行关系。研究多个相关变量间的关系，曾经采用简单相关系数分析。然而这种分析带有很大的片面性。因为已经证明简单相关系数可以剖分为直接作用与间接作用的代数和。一般说来，简单相关系数并不反映一个变量(原因)对另一个变量(结果)的直接作用，因而由简单相关系数分析常会得出错误的结论。人们也试图采用多元线性回归分析法来研究多个相关变量间的关系。这比用简单相关系数分析进了一大步。但多元线性回归分析也有一定的局限性，因为通过偏回归系数的计算仅指出了各原因对结果的直接作用，未能解决两两相关的原因共同对结果作用的问题；且由于偏回归系数带有单位，不能直接由他们比较各原因对结果直接作用的大小。若各自变量间相互独立，则多元线性回归分析是一种十分有效的方法。然而在动、植物的多个数量性状的研究中，各性状间常常是彼此相关的，不具备相互独立的条件。所以采用多元线性回归分析法研究动、植物多个数量性状间的关系有时也达不到预定的目的。由 S.Wright(1921)提出，并经遗传育种工作者不断

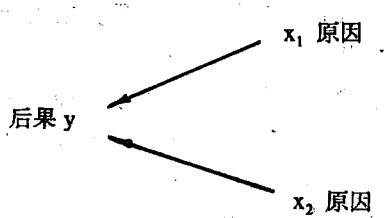
完善与改进的通径分析(Path Analysis)在研究多个相关变量间关系中具有直观、精确等优点，在遗传育种工作中广泛应用于研究遗传相关、近交系数、亲缘系数、遗传力，确定综合选择指数、复合育种值，剖分性状间的相关系数为直接作用与间接作用的代数和，等等。通径分析是遗传育种工作者研究多个相关变量间关系的有力工具。它早已广泛应用于动物育种和遗传研究。近十几年来，在植物育种研究上应用通径分析也日趋广泛。本书将详细阐述通径分析的基本原理与方法并举例说明其应用，以飨读者。

目　录

第一节　通径系数与决定系数

先介绍通径系数与决定系数这二个基本概念，并用通径图表示相关变量间的关系。为了便于理解，先考虑三个相关变量间的情况，然后推广到一般。

设有三个相关变量 y，x_1，x_2，其中 y 是后果(即依变量)，x_1、x_2 是原因(即自变量)。两原因 x_1、x_2 间的关系有两种可能：(1) x_1 与 x_2 相互独立；(2) x_1 与 x_2 彼此相关。当二原因间相互独立时，可用图 1-1 表示如下：



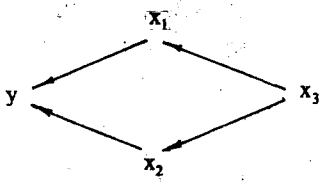
(图 1-1)

当二原因间彼此相关时，二者是平行关系，互为因果或有一个共同原因，可用图 1-2 表示如下：(图见下页)

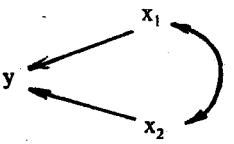
其中 x_3 是 x_1、x_2 的公共原因。若用一条双箭头线"x_1 ⌒ x_2"代替两条单箭头线。"$x_1 \leftarrow x_3 \rightarrow x_2$"，图 1-2 可改画为图 1-3：(图见下页)

图 1-3 中的单箭头"→"表示变量间存在着因果关系，方向由原因到结果，称为通径。图 1-3 中的双箭头"⌒"表示变量间存在着平行关系，互为因果，称为相关线。相关线相当于两条尾端相连的通径。这种用来表示相关变量间因果

关系与平行关系的箭形图叫通径图。



(图 1-2)



(图 1-3)

通过作通径图，形象直观地表达了相关变量间的关系。但这只是定性地表达，还须进一步用数量表示因果关系中原因对结果影响的相对重要性、平行关系中变量间相关的相对重要性。换句话说，还须用数量表示通径图中"通径"的相对重要性和"相关线"的相对重要性。表示通径相对重要性的数量叫通径系数；表示相关线相对重要性的数量的相关系数。生物统计学已给出了计算相关系数的方法，即：若二相关变量 x_1、x_2 有 n 组观测值，则 x_1 与 x_2 的相关系数 r_{12} 的计算公式为：

$$r_{12}=\frac{\sum(x_1-\bar{x}_1)(x_2-\bar{x}_2)}{\sqrt{\sum(x_1-\bar{x}_1)^2\sum(x_2-x_2)^2}}$$

下面我们给出通径系数的确切定义和数学表示式。

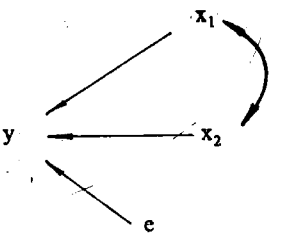
设相关变量 y 与 x_1、x_2 间存在线性关系，回归方程式为：

$$\hat{y}=b_0+b_1x_1+b_2x_2 \qquad (1-1)$$

或

$$y=b_0+b_1x_1+b_2x_2+e \qquad (1-2)$$

其中：y 为依变量；x_1、x_2 为自变量且彼此相关；b_1、b_2 分别为 y 对 x_1、x_2 的偏回归系数；e 为误差项(或剩余项)、相互独立且都服从 $N(O, \sigma_e^2)$。e 与各自变量独立无关①。表示这三个相关变量间关系的通径图见图 1-4。



(图 1-4)

因为偏回归系数带有具体的单位，如 b_1 所带的单位为：y 的单位／x_1 的单位；b_2 所带的单位为：y 的单位／x_2 的单位。一般说来，不能由 b_1、b_2 直接比较原因 x_1、x_2 对结果 y 的影响的重要程度。为了能直接比较各原因对结果影响的重要程度，现先将 y，x_1，x_2 三个变量标准化，约去单位，变为相对数，再研究标准化变量的线性关系。

---

①因为误差项与各自变量独立无关，文献(4)、(5)采用先不考虑误差项的方式引进各原因对结果的通径系数与决定系数的概念，以后再考虑误差项对结果的决定系数与通径系数。那里的作法是正确的。为了进行比较，这里采用同时考虑误差项的方式引入这二个概念。两种方式所得结果完全一致。如果将误差项也当作一个自变量，在前面赋与"偏回归系数 be"并在正规方程中包含 be 的作法显然是错误的，因为这既不符合多元线性回归分析数学模型对误差项的要求，也不符合最小二乘原理。

由(1−2)式可得 y 的平均数$\bar{y}$为:

$$\bar{y} = b_0 + b_1\bar{x}_1 + b_2\bar{x}_2 + \bar{e} \qquad (1-3)$$

将(1−2)式左右两端分别减去(1−3)式的左右两端，得:

$$y - \bar{y} = b_1(x_1 - \bar{x}_1) + b_2(x_2 - \bar{x}_2) + (e - \bar{e}) \qquad (1-4)$$

将(1−4)式两端同除以 y 的标准差 σ_0[①]，且作相应的恒等变换，得:

$$\frac{y-\bar{y}}{\sigma_0} = b_1\frac{\sigma_1}{\sigma_0}\cdot\frac{x_1-\bar{x}_1}{\sigma_1} + b_2\frac{\sigma_2}{\sigma_0}\cdot\frac{x_2-\bar{x}_2}{\sigma_2} + \frac{\sigma_e}{\sigma_0}\cdot\frac{e-\bar{e}}{\sigma_e} \qquad (1-5)$$

其中 σ_1、σ_2、σ_e 分别为 x_1、x_2 与误差项 e 的标准差。记

$$y' = \frac{(y-\bar{y})}{\sigma_0},\quad x'_1 = \frac{(x_1-\bar{x}_1)}{\sigma_1},\quad x'_2 = \frac{(x_2-\bar{x}_2)}{\sigma_2},$$

$$e' = \frac{(e-\bar{e})}{\sigma_e},$$

分别为 y，x_1，x_2，e 的标准化(或标准正态离差)。于是，(1−5)式可改写为:

$$y' = b_1\frac{\sigma_1}{\sigma_0}x'_1 + b_2\frac{\sigma_2}{\sigma_0}x'_2 + \frac{\sigma_e}{\sigma_0}e' \qquad (1-6)$$

或

$$\hat{y} = b_1\frac{\sigma_1}{\sigma_0}x'_1 + b_2\frac{\sigma_2}{\sigma_0}x'_2 \qquad (1-7)$$

---

①确切地说，此处是指y的样本标准差。本书前三节未严格区分总体标准差σ和样本标准差 s 这两个符号，这可从具体的问题中区分开来。

$b_1\frac{\sigma_1}{\sigma_0}$、$b_2\frac{\sigma_2}{\sigma_0}$为变量标准化后的偏回归系数，是不带单位的相对数，可用以直接比较 x'_1，x'_2 对 y′影响的大小，即分别表示了 x_1，x_2 对 y 影响的相对重要性；σ_e/σ_0 则表示了误差项 e 对 y 影响的相对重要性．它们分别为 x_1，x_2，e 到 y 的通径系数．综上所述，我们将通径系数的定义归纳如下:

定义　若相关变量 y，x_1，x_2 间存在线性关系，回归方程式为:

$$\hat{y}=b_0+b_1x_1+b_2x_2$$

或

$$y=b_0+b_1x_1+b_2x_2+e$$

则变量标准化后的各偏回归系数$b_1\frac{\sigma_1}{\sigma_0}$，$b_2\frac{\sigma_2}{\sigma_0}$分别称为原因 x_1、x_2 到结果 y 的通径系数，记为 $P_{0\cdot1}$，$P_{0\cdot2}$；σ_e/σ_0 称为误差项 e 到结果 y 的通径系数，记为 $P_{0\cdot e}$，即

$$P_{0\cdot1}=b_1\frac{\sigma_1}{\sigma_0},\quad P_{0\cdot2}=b_2\frac{\sigma_2}{\sigma_0},\quad P_{0\cdot e}=\frac{\sigma_e}{\sigma_0}$$

通径系数的平方称为决定系数，表示原因(自变量或误差)对结果(依变量)的相对决定程度。x_1，x_2，e 对 y 的决定系数记为 $d_{0\cdot1}$，$d_{0\cdot2}$，$d_{0\cdot e}$，于是

$$d_{0\cdot1}=P^2_{0\cdot1}=(b_1\frac{\sigma_1}{\sigma_0})^2,$$

$$d_{0\cdot2}=P^2_{0\cdot2}=(b_2\frac{\sigma_2}{\sigma_0})^2,$$

$$d_{0\cdot e}=P^2_{0\cdot e}=\frac{\sigma_e^2}{\sigma_0^2}.$$

当相关变量 y、x_1，x_2 间的回归方程式为：

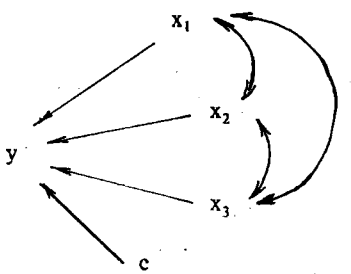
$$\hat{y} = x_1 + x_2,$$

即，$b_0 = 0$，$b_1 = b_2 = 1$，此时通径系数为相应标准差之比，决定系数为相应方差之比①，即

$$P_{0\cdot1} = \frac{\sigma_1}{\sigma_0}, \quad P_{0\cdot2} = \frac{\sigma_2}{\sigma_0};$$

$$d_{0\cdot1} = \frac{\sigma_1^2}{\sigma_0^2}, \quad d_{0\cdot2} = \frac{\sigma_2^2}{\sigma_0^2}.$$

上述定义可以扩展到三个以上的原因(自变量)的情况。例如，若相关变量 y，x_1，x_2，x_3 间存在线性关系，y 为依变量，x_1、x_2、x_3 为自变量，x_1、x_2、x_3 两两相关，通径图如图 1–5 所示。



(图 1–5)

y 与 x_1，x_2，x_3 的回归方程式为：

---

①C、C、Li 在1963出版的《Population Genetics》中把通径系数和决定系数定义为标准差之比和方差之比。C、C、Li 的定义仅是这里所给出的一般定义在偏回归系为 1 时的特殊情况。

$$\hat{y} = b_0 + b_1 x_1 + b_2 x_2 + b_3 x_3$$

或

$$y = b_0 + b_1 x_1 + b_2 x_2 + b_3 x_3 + e$$

则 x_1，x_2，x_3 和 e 到 y 的通径系数分别为:

$$P_{0\cdot1} = b_1 \frac{\sigma_1}{\sigma_0}, \quad P_{0\cdot2} = b_2 \frac{\sigma_2}{\sigma_0},$$

$$P_{0\cdot3} = b_3 \frac{\sigma_3}{\sigma_0}, \quad P_{0\cdot e} = \frac{\sigma_e}{\sigma_0}。$$

x_1，x_2，x_3 和 e 对 y 的决定系数分别为:

$$d_{0\cdot1} = P^2_{0\cdot1} = (b_1 \frac{\sigma_1}{\sigma_0})^2,$$

$$d_{0\cdot2} = P^2_{0\cdot2} = (b_2 \frac{\sigma_2}{\sigma_0})^2,$$

$$d_{0\cdot3} = P^2_{0\cdot3} = (b_3 \frac{\sigma_3}{\sigma_0})^2,$$

$$d_{0\cdot e} = P^2_{0\cdot e} = \frac{\sigma_e^2}{\sigma_0^2}。$$

当 $\hat{y} = x_1 + x_2 + x_3$ 时，即 $b_0=0$，$b_1=b_2=b_3=1$，此时有

$$P_{0\cdot1} = \frac{\sigma_1}{\sigma_0}, \quad P_{0\cdot2} = \frac{\sigma_2}{\sigma_0}, \quad P_{0\cdot3} = \frac{\sigma_3}{\sigma_0};$$

$$d_{0\cdot1} = \frac{\sigma_1^2}{\sigma_0^2}, \quad d_{0\cdot2} = \frac{\sigma_2^2}{\sigma_0^2}, \quad d_{0\cdot3} = \frac{\sigma_3^2}{\sigma_0^2}。$$

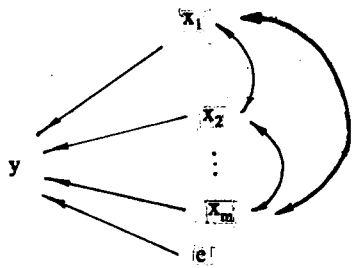
一般，若相关变量 y，x_1，x_2，…，x_m 间存在线性关系，y 为依变量，x_1、x_2、…、x_m 为自变量且两两相关，通径图如图 1-6 所示。(图见下页)

回归方程式为:

$$\hat{y} = b_0 + b_1 x_1 + b_2 x_2 + \cdots + b_m x_m$$

或

$$y = b_0 + b_1 x_1 + b_2 x_2 + \cdots + b_m x_m + e$$



(图 1–6)

则 x_i ($i=1, 2, \cdots, m$) 与 e 到 y 的通径系数为:

$$P_{0\cdot i} = b_i \frac{\sigma_i}{\sigma_0}, \quad (i = 1, 2, \cdots, m);$$

$$P_{0\cdot e} = \frac{\sigma_e}{\sigma_0};$$

x_i ($i=1, 2, \cdots, m$) 与 e 对 y 的决定数为:

$$d_{0\cdot i} = P_{0\cdot i}^2 = (b_i \frac{\sigma_i}{\sigma_0})^2, \quad (i = 1, 2, \cdots, m);$$

$$d_{0\cdot e} = P_{0\cdot e}^2 = \frac{\sigma_e^2}{\sigma_0^2}.$$

特别地，若 $\hat{y} = x_1 + x_2 + \cdots + x_m$，即 $b_0 = 0$，$b_1 = b_2 = \cdots = b_m = 1$，则

$$P_{0\cdot i} = \frac{\sigma_i}{\sigma_0}, \quad d_{0\cdot i} = \frac{\sigma_i^2}{\sigma_0^2} \quad (i = 1, 2, \cdots, m).$$

为了进一步了解通径系数的特性，我们再讨论仅有二个相关变量 y 与 x_1 的情况．如果 y 与 x_1 存在线性关系，回归方程式为：

$$\hat{y} = b_0 + b_1 x_1$$

此时，$P_{0\cdot1} = b_1 \frac{\sigma_1}{\sigma_0}$。而回归系数

$$b_1 = \frac{\sum(x_1 - \bar{x}_1)(y - \bar{y})}{\sum(x_1 - \bar{x}_1)^2},$$

所以

$$\begin{aligned} P_{0\cdot1} &= \frac{\sum(x_1 - \bar{x}_1)(y - \bar{y})}{\sum(x_1 - \bar{x}_1)^2} \cdot \frac{\sigma_1}{\sigma_0} \\ &= \frac{\sum(x_1 - \bar{x}_1)(y - \bar{y})}{\sum(x_1 - \bar{x}_1)^2} \cdot \frac{\sqrt{\sum(x_1 - \bar{x}_1)^2/(n-1)}}{\sqrt{\sum(y - \bar{y})^2/(n-1)}} \\ &= \frac{\sum(x_1 - \bar{x}_1)(y - \bar{y})}{\sqrt{\sum(x_1 - \bar{x}_1)^2 \sum(y - \bar{y})^2}} \\ &= r_{10} \quad (r_{10} \text{ 表示 } x_1 \text{ 与 y 的相关系数}). \end{aligned}$$

表明，在一元线性回归分析中，原因 x_1 到结果 y 的通径系数 $P_{0\cdot1}$ 就等于 x_1 与 y 的相关系数 r_{10}（在下一节中将证明，在一定条件下，这个结论对于多元线性回归分析也成立）。此时，通径系数与相关系数在数量上相等，都是不带单位的相对数，但两者之间却有实质上的区别：通径系数表示的是相关变量间的因果关系，是有方向的；而相关系数表示的是相关变量间的平行关系，互为因果，是没有方向的。可以说通径系数是自变量与依变量间有方向的相关系数。

综上所述，关于通径系数我们有如下结论：