

《生物实验设计》

第十二章 逐步回归与通径分析

王超

广东药科大学

Email: wangchao@gdpu.edu.cn

2022-10-24



廣東藥科學大學
GUANGDONG PHARMACEUTICAL UNIVERSITY

第十二章 逐步回归与通径分析

一元与多元回归

- 对于多变量资料，既包含对因变量 y 具有显著线性效应的自变量，又包含对 y 不具有显著线性效应的自变量
- 分析中必须将不具有显著效应的自变量予以舍去，使所得到的多元线性回归方程中的自变量对因变量 y 均具有显著效应，这就是最优多元线性回归方程
- 可通过逐步回归方法建立最优回归方程，来简洁准确地分析和预测因变量 y 的反应
- 通径分析是另一种研究多个相关变量间线性关系的统计方法

第一节 逐步回归分析

- 逐步回归分析的两种基本途径：

- 向前逐步回归

- 从一元回归分析开始，按各自变量对 y 作用的秩次，依次每步仅选入一个对 y 作用显著的自变量
 - 每引入一个自变量后，对在此之前已引入的自变量进行重新检验，有不显著者即舍弃
 - 直到选入的自变量都显著，未被选入的自变量都不显著为止

- 向后逐步回归

- 从 m 元回归分析开始，每步舍去一个不显著且偏回归平方和为最小的自变量
 - 每次社区一个偏回归不显著且平方和最小的自变量之后，需对回归方程和各自变量重新进行假设检验
 - 直到回归方程所包含的自变量全部显著
 - 自变量个数较少，且大多都显著时，这种方法就比较实用

第一节 逐步回归分析

一、逐个淘汰不显著自变量的回归方法

- m 元回归分析
 - 若各自变量的偏回归皆显著，分析结束
 - 若有一个或一个以上自变量的偏回归不显著，则舍弃偏回归平方和最小的自变量，进入下一步
- $m - 1$ 元回归分析
 - 将舍弃的自变量所在的行、列及其 K 列划去，重新计算 $m - 1$ 阶系数矩阵的逆矩阵元素
 - 如果仍有自变量偏回归不显著，则再将偏回归平方和最小的自变量舍去，进入下一步
- 重复进行，直至留下所有自变量的偏回归系数皆显著，即得到最优回归方程

第一节 逐步回归分析

一、逐个淘汰不显著自变量的回归方法

第一节 逐步回归分析 一、逐个淘汰不显著自变量的回归方法

在进行 m 元回归分析的基础上，余下自变量的偏回归系数和逆矩阵 A^{-1} 中 c_{ij} 的计算，可根据舍弃前的偏回归系数和 c_{ij} ，通过公式直接求出

设 x_k 为舍弃的自变量，则

$$b_i^* = b_i - \frac{c_{ik}b_k}{c_{kk}} (i \neq k)$$
$$c_{ij}^* = c_{ij} - \frac{c_{ik}c_{kj}}{c_{kk}} (i, j \neq k)$$

第一节 逐步回归分析 一、逐个选入显著自变量的回归方法

- 每一次都选入一个显著的自变量，其方法步骤如下
 - 计算各变量的简单相关系数，得 $m+1$ 阶相关矩阵 $R^{(0)}$

$$R^{(0)} = \begin{bmatrix} r_{11}^{(0)} & r_{12}^{(0)} & \cdots & r_{1m}^{(0)} & r_{1y}^{(0)} \\ r_{21}^{(0)} & r_{22}^{(0)} & \cdots & r_{2m}^{(0)} & r_{2y}^{(0)} \\ \vdots & & & & \\ r_{m1}^{(0)} & r_{m2}^{(0)} & \cdots & r_{mm}^{(0)} & r_{my}^{(0)} \\ r_{y1}^{(0)} & r_{y2}^{(0)} & \cdots & r_{ym}^{(0)} & r_{yy}^{(0)} \end{bmatrix}$$

- 简记为 $R^{(0)} = (r_{ij}^{(0)}), (i, j = 1, 2, 3, \dots, m, y)$

第一节 逐步回归分析

二、逐个选入显著自变量的回归方法

- 选入自变量逐步回归

- 以 $R^{(0)}$ 为基础，每进行一步回归选入一个显著的自变量，并对相关矩阵做一次变换
- 在第一步，将 $R^{(0)} = (r_{ij}^{(0)})$ 变为 $R^{(1)} = (r_{ij}^{(1)})$
- 在第二步，将 $R^{(1)} = (r_{ij}^{(1)})$ 变为 $R^{(2)} = (r_{ij}^{(2)})$
- 在第 k 步，将 $R^{(k-1)} = (r_{ij}^{(k-1)})$ 变为 $R^{(k)} = (r_{ij}^{(k)})$

第一节 逐步回归分析

二、逐个选入显著自变量的回归方法

在第 k 步 ($k = 1, 2, \dots, m + 1$), 由下式算得任一尚未入选自变量 x_i 的标准偏回归平方和

$$U_i^{(k)} = \frac{(r_{iy}^{(k-1)})^2}{r_{ii}^{(k-1)}}$$

设最大 $U_i^{(k)}$ 的自变量 $x_i (i = l)$, 则 x_l 在第 k 步是否入选由下式决定

$$F = \frac{U_l^{(k)}}{\frac{r_{yy}^{(k-1)} - U_l^{(k)}}{n - m - 1}}$$

第一节 逐步回归分析 二、逐个选入显著自变量的回归方法

若 $F > F_\alpha$ ，则引入自变量 x_l ，并将 $R^{(k-1)}$ 变换成 R^k 。变换时由元素 $r_{ij}^{(k-1)}$ 计算元素 $r_{ij}^{(k)}$ 的通式为

$$\begin{cases} r_{ll}^{(k)} = \frac{1}{r_{ll}^{(k-1)}} \\ r_{lj}^{(k)} = \frac{r_{lj}^{(k-1)}}{r_{ll}^{(k-1)}} \\ r_{il}^{(k)} = -\frac{r_{il}^{(k-1)}}{r_{ll}^{(k-1)}} \\ r_{ij}^{(k)} = r_{ij}^{(k-1)} - \left(\frac{r_{il}^{(k-1)} r_{lj}^{(k-1)}}{r_{ll}^{(k-1)}} \right) \end{cases}$$

第二节 通径分析

- 由于各个 x_i 单位不同和 x_i 变异度不同, 各个 x_i 对 y 的贡献大小就不能直接进行比较
- 相关分析中, 变量之间是一种平等关系, x_i 与 y 仅表示两个变量之间的密切程度
- 通过通径分析, 可将相关系数 r_{ij} 剖分为 x_i 对 y 的直接作用和 x_i 通过与其相关的各个 x_i 对 y 的间接作用
- 通径分析是分析相关变量间因果关系的一种统计方法