通径分析

《美国数学建模竞赛》

完整课程请长按下方二维码

目录页

目录
CONTENTS

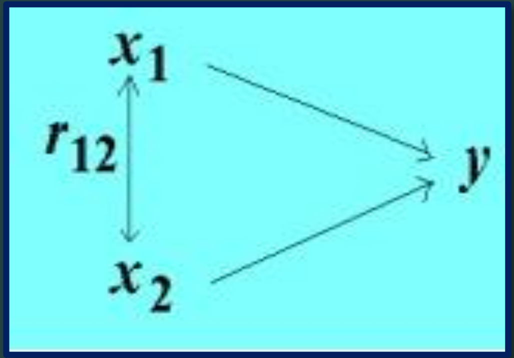
方法使用的背景

- 仅仅研究两个变量的关系　简单相关系数
- 多个相关变量中研究两个变量之间的关系　偏相关系数
- 多个不相关自变量与一个因变量之间的关系　多元回归
- 多个相关的自变量与一个因变量之间的关系　通径分析
- 多个相关的因变量和多个相关的自变量之间的关系

典型相关系数

通径分析数学原理

- 图中 $\rightarrow$ 表示自变量间存在因果关系，方向由原因到结果，称为通径。
- $\leftrightarrow$ 表示变量间存在平行关系，称为相关线。



x_i 指向 y 的连接线为 $x_i \rightarrow y$ 直接通径。

在直接通径上，若 x_i 的取值增加一个标准差单位时, y 将要改变的标准差单位数 P_{iy} 称为通径 $x_i \rightarrow y$ 的系数。

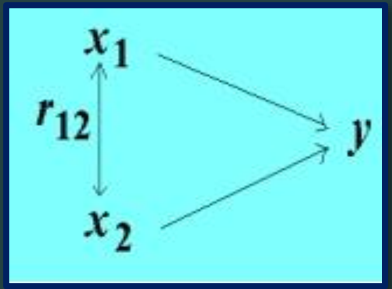
类似于概率论中的概率的乘法公式

$$P(AB) = P(A)P(B/A)$$

间接通径 $x_i \to x_j \to y$ 的系数为 $p_{i\to j\to y} = r_{ij}\, p_{y\cdot i}$

间接通径 $x_j \to x_i \to y$ 的系数为 $p_{j\to i\to y} = r_{ij}\, p_{y\cdot j}$

$$\begin{cases} p_{1y} + r_{12}p_{2y} = r_{1y} \\ r_{21}p_{1y} + p_{2y} = r_{2y} \end{cases}$$



- 推广到一般，即一个依变量 y 与 m 个自变量的情形，则有通径方程：

$$\begin{cases} p_{1y} + r_{12} p_{2y} + r_{13} p_{3y} + \cdots + r_{1m} p_{my} = r_{1y} \\ r_{21} p_{1y} + p_{2y} + r_{23} p_{3y} + \cdots + r_{2m} p_{my} = r_{2y} \\ r_{31} p_{1y} + r_{32} p_{2y} + P_{3y} + \cdots + r_{3m} p_{my} = r_{3y} \\ \cdots\cdots \\ r_{m1} p_{1y} + r_{m2} p_{2y} + r_{m3} p_{3y} + \cdots + p_{my} = r_{my} \end{cases}$$

已知相关系数 r_{ij}, r_{iy}，求 P_{iy}，这实际上就是个解线性方程组的过程

计算步骤

- 1. 计算所有自变量与因变量的简单相关系数，并做相关性检验，排除与因变量不相关的自变量。
- 2. 计算余下所有自变量之间的相关系数。
- 3. 建立通径方程。
- 4. 解方程组，计算出直接通径系数

例　假设得到的简单相关系数分解如下：

$$\begin{cases} p_1 - 0.135742\, p_2 + 0.5007305\, p_3 = 0.8973138 \\ -0.135742\, p_1 + p_2 - 0.148887\, p_3 = 0.0461919 \\ 0.5007305\, p_1 - 0.148887\, p_2 + p_3 = 0.6889796, \end{cases}$$

则直接通径系数为：

$$p_1 = 0.753,\quad p_2 = 0.199,\quad p_3 = 341$$

总结与体会

- 从回归的角度解释，因变量与若干个自变量拟合回归方程，自变量前的系数称为偏回归系数；若将因变量与若干个自变量都标准化后再拟合回归方程，自变量前的系数就是通径系数。
- 具体解释可参考教材《数学建模方法入门及其应用》，科学出版社，2018.4