

10 多元综合评价分析方法

多元统计分析的应用范围很广，无论自然科学还是社会科学，无论是开发研究还是实际预测和决策，都有许多成功的案例。本章简要介绍多元统计分析的相关基础和常用的多元统计分析评价基本方法。

10.1 方差分析

在评价系统研究中，我们所关心的参数往往是受多因素影响的，如作物产量的高低、城市空气质量的好坏等。如果要提高产量或者要改善环境条件，就必须对影响生产或环境的多因素进行分析，确定哪些因素是主要的，以及将这些因素控制在哪个水平或位级上就能使农作物产量提高或环境质量改善。

作为影响研究参数结果的因素，有些是可控的，这些可控制因素一般可以在有控制的重复该试验或观测所获得的数据中得到反映。方差分析就是通过对试验或观测所获得的数据之间的差异，分析推断各影响因素所起作用的一种数理统计方法。英国统计学家费歇尔在20世纪20年代首先把方差分析应用到农业试验中。

这里仅介绍单因素和双因素的方差分析，主要讨论方差分析的一些具体算法，对涉及概念、计算公式的数学原理只作简单介绍。

10.1.1 单因素方差分析

A 单因素方差分析模型

设影响研究结果的因素只有一个，令其为 A 。对 A 取 k 个不同水平（或称状态），分别进行了 $n_1, \dots, n_k$ 次重复试验（观测），共得到 $N = \sum_{i=1}^k n_i$ 个观测数据 x_{ij} ，

$i=1,2,\dots,k$ ； $j=1,2,\dots,n_i$ 。试验（观测）的目的是要分析因素 A 的不同水平对

试验所研究的结果是否有显著的影响。假设 x_{ij} 有如下的数据结构式：

$$x_{ij} = \mu + \alpha_i + \varepsilon_{ij} \quad (i=1,2,\dots,k; j=1,2,\dots,n_i) \quad (10-1)$$

式中， μ 表示总体均值； α_i 表示水平 A_i 对观测数据的影响，即水平 A_i 的效应； ε_{ij} 表示第 i 个水平在第 j 次试验（观测）中的随机误差。假定 ε_{ij} 相互独立，且 $\varepsilon_{ij} \sim N(0, \delta^2)$ 。

于是，根据模型式（10-1），上述试验（观测）的目的可归结为检验假设：

$$H_0: \alpha_1 = \alpha_2 = \dots = \alpha_k = 0 \quad (10-2)$$

B 方差分析原理

基本思想是：根据所设的数据结构将总偏差平方和分解为各个不同水平效应的偏差平方和及随机误差所引起的偏差平方和，然后利用正态分布假设，构造 F 统计量进行显著性检验。

令： $\bar{x}_i = \frac{1}{n_i} \sum_{j=1}^{n_i} x_{ij}$ 表示水平 A_i 的样品平均值；

$\bar{\bar{x}} = \frac{1}{N} \sum_{i=1}^k \sum_{j=1}^{n_i} x_{ij}$ 表示全部观测值的总平均。

总偏差平方和可分解为：

$$\begin{aligned} S_T &= \sum_{i=1} \sum_{j=1} (x_{ij} - \bar{x})^2 = \sum_{i=1} n_i (\bar{x}_{i\cdot} - \bar{x})^2 + \sum_{i=1} \sum_{j=1} (x_{ij} - \bar{x}_{i\cdot})^2 \\ &= S_E + S_A \end{aligned}$$

(10-3)

式中， $S_E = \sum_{i=1} \sum_{j=1} (x_{ij} - \bar{x}_{i\cdot})^2 = \sum_{i=1} \sum_{j=1} (\varepsilon_{ij} - \bar{\varepsilon}_i)^2$ 表示各组内随机因素而产生的偏差平方和，简称组内平方和（或误差平方和）。

$$S_A = \sum_{i=1} \sum_{j=1} (\bar{x}_{i\cdot} - \bar{x})^2 = \sum_{i=1} n_i (\bar{x}_{i\cdot} - \bar{x})^2 = \sum_{i=1} n_i (\alpha_i + \bar{\varepsilon}_i - \bar{\varepsilon})^2$$

(10-4)

表示由各组不同水平之间的差异而产生的偏差平方和，简称组间平方和。

在上述关于模型式（10-1）的假设下，统计量 S_T/δ^2 ， S_E/δ^2 ， S_A/δ^2 分别服从自由度为 $N-1$ ， $N-k$ ， $k-1$ 的 χ^2 分布，且 S_E 与 S_A 相互独立，故统计量：

$$F = \frac{S_A/\delta^2 (k-1)}{S_E/\delta^2 (N-k)} \sim F(k-1, N-k)$$

(10-5)

可用来检验假设 H_0 。对于给定的显著性水平 α ，若计算值 $F > F_\alpha(k-1, N-k)$ ，则拒绝 H_0 ，即认为因素 A 的不同水平对所研究的指标值的影响有显著差异。一般，当 $\alpha=0.05$ 时， $F > F_\alpha$ 表示水平效应显著，记作 F^* ；当 $\alpha=0.01$ 时， $F > F_\alpha$ 表示水平效应高度显著，记作 F^{**} 。

C 单因素方差分析计算步骤

方差分析的计算和检验程序可按下列方差分析表 10-1 进行。

- (1) 计算样品（观测值）均值 $\bar{x}_{i\cdot}$ 及总均值 $\bar{x}$ ；
- (2) 计算组间平方和 S_A 及误差平方和 S_E ；
- (3) 计算均方 $\overline{S_A} = S_A/(k-1)$ ， $\overline{S_E} = S_E/(N-k)$ 及 $F = \overline{S_A}/\overline{S_E}$ 。

表 10-1 单因素方差分析表

方差来源	平方和	自由度	均方	F
组 间	$S_A = \sum_{i=1} n_i (\bar{x}_{i\cdot} - \bar{x})^2$	$k-1$	$\overline{S_A} = S_A/(k-1)$	$\frac{\overline{S_A}}{\overline{S_E}}$

误 差	$S_E = \sum_{i=1} \sum_{j=1} (x_{ij} - \bar{x}_i)^2$	$N - k$	$\overline{S_E} = S_E / (N - k)$	
总 合	$S_T = S_E + S_A$	$N - 1$		

D 算例

为了考察用来处理水稻种子的 4 种不同药剂对水稻生长的影响，选择一块各种条件（如土质、气候、管理）基本均匀的土地，将其分成 16 块。在每 4 块试验地里种下用同一种药剂处理过的水稻种子。试验结果——苗高见表 10-2。

试验中唯一的可控因素——药剂用 A 表示，用 4 种不同的药剂处理称为 A 的 4 个不同水平，分别记为 A_1, A_2, A_3, A_4 。这里每个水平的试验（观测）次数均相同， $k = 4, N = 16$ 。其参数计算见表 10-2，方差分析见表 10-3。

表 10-2 不同药剂处理的苗高

药剂 A 重复试验次数	不同水平下的苗高 x				参数
	A_1	A_2	A_3	A_4	
1	23	21	20	22	
2	19	24	18	25	
3	21	27	19	27	
4	13	20	15	26	
总和 T_i	76	92	72	100	$T = 340$
平均 $\bar{x}_i$	19	23	18	25	$\bar{x}_i = 21.25$

表 10-3 不同药剂处理的水稻苗高的方差计算表

方差来源	平方和	自由度	均方和	F
组间（因素 A）	131	3	43.34	4.60
组内（误差）	114	12	9.5	
总 和	245	15		

设显著水平 $\alpha = 0.05$ ，查 F 分布临界值表有：

$F_{\alpha}(k-1, N-k) = F_{0.05}(4-1, 16-4) = 3.49$

由于 $4.60 > 3.49$ ，故拒绝 H_0 ，即不同药剂对水稻的生长有显著影响。同样查表有

$F_{0.01}(4-1, 16-4) = 5.95$ ，则 $F < F_{0.01}$ ，表示不同药剂对水稻的生长不存在高度显著。

10.1.2 多因素方差分析

A 无交互作用的方差分析

实际问题中常会遇到两个或两个以上因素同时影响试验（观测）结果的情形。通过对两个因素的方差分析，能确定两个因素各自的主效应及它们的交互效应；通过对 3 个因素的方差分析，能确定 3 个因素各自的主效应、两两交互效应及 3 个因素的交互效应等。由于多因

以上内容仅为本文档的试下载部分，为可阅读页数的一半内容。如要下载或阅读全文，请访问：<https://d.book118.com/246240043121010031>