附录 数理统计和线性代数基础

第六节 线性模型和方差分析

这里用两因素（用 A 和 B 表示）的因子设计说明方差分析中的线性模型，假定因素 A 有 m 个水平，因素 B 有 n 个水平，y_{ijk} 表示因子组合 $\mathrm{A}_i\mathrm{B}_j$ 下的第 k 个观测值。

一、线性模型的建立

用 μ_{ij} 表示因子组合 $\mathrm{A}_i\mathrm{B}_j$ 的理论值，则观测值 y_{ijk} 可先分解为

$$y_{ijk} = \mu_{ij} + \varepsilon_{ijk}$$

其中 ε_{ijk} 为试验误差，相互间独立，且服从均值为 0、方差为 σ_ε^2 的正态分布，即 $\varepsilon_{ijk} \sim N(0,\sigma_\varepsilon^2)$ 。进一步，设定：

$$\mu = \frac{1}{mn}\sum_{i=1}^{m}\sum_{j=1}^{n}\mu_{ij}\ ,\quad \mu_{i\bullet} = \frac{1}{n}\sum_{j=1}^{n}\mu_{ij}\ ,\quad \mu_{\bullet j} = \frac{1}{m}\sum_{i=1}^{m}\mu_{ij}$$

可以看出，μ 即所有组合的平均数，$\mu_{i\bullet}$ 为因素 A 的第 i 个水平的平均数，$\mu_{\bullet j}$ 为因素 B 的第 j 个水平的平均数。设定因子水平 A_i 的效应值为 $a_i = \mu_{i\bullet} - \mu$，$\mathrm{B}_j$ 的效应值为 $b_j = \mu_{\bullet j} - \mu$，显然因子 A 的所有效应值之和为 0，因子 B 的所有效应值之和也为 0，因子组合 $\mathrm{A}_i\mathrm{B}_j$ 的理论值 μ_{ij} 可进一步分解为

$$\mu_{ij} = \mu + a_i + b_j + [(\mu_{ij} - \mu) - a_i - b_j]$$

上式中，$(\mu_{ij} - \mu)$ 为组合 $\mathrm{A}_i\mathrm{B}_j$ 的效应值，因此 $[(\mu_{ij} - \mu) - a_i - b_j]$ 为组合 $\mathrm{A}_i\mathrm{B}_j$ 的效应值减去 A_i 和 B_j 的效应值，它衡量的是 A_i 和 B_j 搭配时的互作效应，用符号 $(ab)_{ij}$ 表示。容易验证

$$\sum_{i=1}^{m}(ab)_{ij}=\sum_{j=1}^{n}(ab)_{ij}=0$$

根据试验材料的不同，模型[A-11]的效应可以是固定的，称为固定模型（Fixed model）；也可以是随机的，称为随机模型（Random model）；当然还会出现有些效应是固定效应，有些效应是随机效应的情形，称为混和模型（Mixed model）。判断模型中的效应应该视为固定的还是随机的，可借助以下两个原则。

（1）当因子的水平是完全可以控制的时候，因子效应视为固定效应；当因子的水平不是完全可以控制的时候，因子效应视为随机效应。

（2）当试验个体是人为指定的时候，我们不一定要把结论推广到试验群体以外的其它群体，这时试验个体的效应为固定效应；当试验个体是随机挑选的一个样本的时候，我们希望从样本推断总体的性质，试验个体的效应为随机效应。

这样就得到固定效应线性模型的完整表述：

$$y_{ijk}=\mu+a_i+b_j+(ab)_{ij}+\varepsilon_{ijk} \qquad \text{[A-19]}$$

$$\sum_{i=1}^{m}a_i=\sum_{j=1}^{n}b_j=0,\quad \sum_{i=1}^{m}(ab)_{ij}=\sum_{j=1}^{n}(ab)_{ij}=0$$

$$a_i\sim(0,\sigma_{\mathrm{A}}^2),\quad b_j\sim(0,\sigma_{\mathrm{B}}^2),\quad (ab)_{ij}\sim(0,\sigma_{\mathrm{AB}}^2),\quad \varepsilon_{ijk}\sim N(0,\sigma_{\varepsilon}^2)$$

其中$i=1,\cdots,m$表示因子 A 的第 i 个水平，$j=1,\cdots,n$表示因子 B 的第 j 个水平，$k=1,\cdots,r$表示第 k 个重复。

在随机模型中，各效应不再是一个数值，而是一个随机变量。随机效应线性模型的完整表述为：

$$y_{ijk}=\mu+a_i+b_j+(ab)_{ij}+\varepsilon_{ijk} \qquad \text{[A-20]}$$

$$a_i \sim (0,\sigma_{\mathrm{A}}^2)，\ b_j \sim (0,\sigma_{\mathrm{B}}^2)，\ (ab)_{ij} \sim (0,\sigma_{\mathrm{AB}}^2)，\ \varepsilon_{ijk} \sim N(0,\sigma_{\varepsilon}^2)$$

二、线性模型的方差分析

固定模型[A-11]和随机模型[A-12]有相同的自由度分解和平方和的分解，因此也有相同的均方，所不同的是期望均方这一项（表 A-5）。表 A-5 中平方和的计算如下。设：

$$\bar{y} = \tfrac{1}{mnr}\sum_{i=1}^{m}\sum_{j=1}^{n}\sum_{k=1}^{r} y_{ijk}，\quad \bar{y}_{ij\bullet} = \tfrac{1}{r}\sum_{k=1}^{r} y_{ijk}，$$

$$\bar{y}_{i\bullet\bullet} = \tfrac{1}{nr}\sum_{j=1}^{n}\sum_{k=1}^{r} y_{ijk}，\quad \bar{y}_{\bullet j\bullet} = \tfrac{1}{mr}\sum_{i=1}^{m}\sum_{k=1}^{r} y_{ijk}$$

这样，我们可以将观测值与总平均数的离差分解为：

$$y_{ijk} - \bar{y} = (\bar{y}_{i\bullet\bullet} - \bar{y}) + (\bar{y}_{\bullet j\bullet} - \bar{y}) + (\bar{y}_{ij\bullet} - \bar{y}_{i\bullet\bullet} - \bar{y}_{\bullet j\bullet} + \bar{y}) + (y_{ijk} - \bar{y}_{ij\bullet})$$

则总平方和可分解为

$$SS_{\mathrm{T}} = \sum_{i=1}^{m}\sum_{j=1}^{n}\sum_{k=1}^{r}(y_{ijk} - \bar{y})^2 = SS_{\mathrm{A}} + SS_{\mathrm{B}} + SS_{\mathrm{AB}} + SS_{\mathrm{E}}$$

其中

$$SS_{\mathrm{A}} = nr\sum_{i=1}^{m}(\bar{y}_{i\bullet\bullet} - \bar{y})^2，$$

$$SS_{\mathrm{B}} = mr\sum_{j=1}^{n}(\bar{y}_{\bullet j\bullet} - \bar{y})^2，$$

$$SS_{\mathrm{AB}} = r\sum_{i=1}^{m}\sum_{j=1}^{n}(\bar{y}_{ij\bullet} - \bar{y}_{i\bullet\bullet} - \bar{y}_{\bullet j\bullet} + \bar{y})^2，$$

$SS_{\varepsilon} = SS_{\mathrm{T}} - SS_{\mathrm{A}} - SS_{\mathrm{B}} - SS_{\mathrm{AB}}$ 或者

$$SS_{\varepsilon}=\sum_{i=1}^{m}\sum_{j=1}^{n}\sum_{k=1}^{r}(y_{ijk}-\bar{y}_{ij\bullet})^2$$

表 A-5 中均方由平方和除以自由度获得。根据表 A-5 中的期望均方，对固定模型来说，各方差的估计量分别为：

$$\sigma_{\mathrm{A}}^2=\tfrac{1}{nr}(MS_{\mathrm{A}}-MS_{\varepsilon})$$

$$\sigma_{\mathrm{B}}^2=\tfrac{1}{mr}(MS_{\mathrm{B}}-MS_{\varepsilon})$$

$$\sigma_{\mathrm{AB}}^2=\tfrac{1}{r}(MS_{\mathrm{AB}}-MS_{\varepsilon})$$

$$\sigma_{\varepsilon}^2=MS_{\varepsilon}$$

方差σ_{A}^2、σ_{B}^2和σ_{AB}^2显著性检验的 F 统计量分别为：

$$F_{\mathrm{A}}=\frac{MS_{\mathrm{A}}}{MS_{\varepsilon}}\sim F[(m-1),mn(r-1)]$$

$$F_{\mathrm{B}}=\frac{MS_{\mathrm{B}}}{MS_{\varepsilon}}\sim F[(n-1),mn(r-1)]$$

$$F_{\mathrm{AB}}=\frac{MS_{\mathrm{AB}}}{MS_{\varepsilon}}\sim F[(m-1)(n-1),mn(r-1)]$$

表 A-5 两因子试验设计的方差分析

方差来源	自由度	平方和	均方	期望均方	
				固定模型	随机模型
因子 A	m-1	SS_{A}	$MS_{\mathrm{A}}= SS_{\mathrm{A}}/(m-1)$	$nr\sigma_{\mathrm{A}}^2+\sigma_{\varepsilon}^2$	$nr\sigma_{\mathrm{A}}^2+r\sigma_{\mathrm{AB}}^2+\sigma_{\varepsilon}^2$
因子 B	n-1	SS_{B}	$MS_{\mathrm{B}}=SS_{\mathrm{B}}/(n-1)$	$mr\sigma_{\mathrm{B}}^2+\sigma_{\varepsilon}^2$	$mr\sigma_{\mathrm{B}}^2+r\sigma_{\mathrm{AB}}^2+\sigma_{\varepsilon}^2$

A×B	$(m\text{-}1)(n\text{-}1)$	SS_{AB}	$MS_{AB}=SS_{AB}/[(m\text{-}1)(n\text{-}1)]$	$r\sigma_{AB}^2+\sigma_\varepsilon^2$	$r\sigma_{AB}^2+\sigma_\varepsilon^2$
随机误差	$mn(r\text{-}1)$	SS_ε	$MS_\varepsilon=SS_\varepsilon/[mn(r\text{-}1)]$	σ_ε^2	σ_ε^2

对随机模型来说，各方差的估计量分别为：

$$\sigma_{\mathrm{A}}^2=\tfrac{1}{nr}(MS_{\mathrm{A}}-MS_{\mathrm{AB}})$$

$$\sigma_{\mathrm{B}}^2=\tfrac{1}{mr}(MS_{\mathrm{B}}-MS_{\mathrm{AB}})$$

$$\sigma_{\mathrm{AB}}^2=\tfrac{1}{r}(MS_{\mathrm{AB}}-MS_{\mathrm{E}})$$

$$\sigma_\varepsilon^2=MS_\varepsilon$$

方差σ_{A}^2、σ_{B}^2和σ_{AB}^2显著性检验的F统计量分别为：

$$F_{\mathrm{A}}=\frac{MS_{\mathrm{A}}}{MS_{\mathrm{AB}}}\sim F[(m-1),(m-1)(n-1)]$$

$$F_{\mathrm{B}}=\frac{MS_{\mathrm{B}}}{MS_{\mathrm{AB}}}\sim F[(n-1),(m-1)(n-1)]$$

$$F_{\mathrm{AB}}=\frac{MS_{\mathrm{AB}}}{MS_\varepsilon}\sim F[(m-1)(n-1),mn(r-1)]$$

其它试验设计的线型模型和方程分析表可参照上述过程去建立。