

§ 7.4 两因素方差分析

设在某试验中有两个因素都在变化, 因素 A 取 r 个不同水平 A_1, A_2, \dots, A_r . 因素 B 取 s 个不同水平 B_1, B_2, \dots, B_s . 因素组合 $A_i B_j$ 条件下的试验结果 X_{ij} 相互独立且服从分布 $N(\mu_{ij}, \sigma^2)$. 记

$$\left\{ \begin{array}{l} \mu = \frac{1}{rs} \sum_{i=1}^r \sum_{j=1}^s \mu_{ij} \\ \mu_{i\cdot} = \frac{1}{s} \sum_{j=1}^s \mu_{ij} \quad i = 1, \dots, r \quad \left\{ \begin{array}{l} \alpha_i = \mu_{i\cdot} - \mu \quad (i = 1, \dots, r) \\ \beta_j = \mu_{\cdot j} - \mu \quad (j = 1, \dots, s) \end{array} \right. \\ \mu_{\cdot j} = \frac{1}{r} \sum_{i=1}^r \mu_{ij} \quad j = 1, \dots, s \end{array} \right.$$

称 μ 为一般平均，它是全部 $r \times s$ 个总体均值的平均数. $\mu_{i\cdot}$ 是 s 个总体 $X_{i1}, X_{i2}, \dots, X_{is}$ 均值的平均数, 试验条件为

$$A_i B_1, A_i B_2, \dots, A_i B_s$$

$\mu_{k\cdot}$ 是 s 个总体 X_{k1}, \dots, X_{ks} 均值的平均数, 试验条件为

$$A_k B_1, A_k B_2, \dots, A_k B_s$$

它们的不同是由于因素 A 取不同水平造成的，因此 α_i 是因素 A 的水平 A_i 带来的影响，称为因素 A 的第 i 个水平的效应，同样 β_j 称为因素 B 的第 j 个水平

的效应. $\sum_{i=1}^r \alpha_i = 0, \quad \sum_{j=1}^s \beta_j = 0.$

$\mu_{ij} - \mu$ 反映了因素水平组合 $A_i B_j$ 的总效应，总效应不一定等于 A_i 的效应 α_i 与 B_j 的效应 β_j 之和. 它们的关系有以下两种情况

(1) 对所有的 i 与 j 有

$$\mu_{ij} - \mu = \alpha_i + \beta_j \quad i = 1, 2, \dots, r. \quad j = 1, 2, \dots, s.$$

此时, $\mu_{ij} = \mu + \alpha_i + \beta_j$. 这时称为无交互作用的双因素方差分析模型.

(2) 对某些 i 与 j 有

$$\mu_{ij} - \mu \neq \alpha_i + \beta_j \quad i = 1, 2, \dots, r. \quad j = 1, 2, \dots, s.$$

这时称为有交无作用的双因素方差分析模型.

$$\text{记} \quad (\alpha\beta)_{ij} = \mu_{ij} - \mu - \alpha_i - \beta_j$$

称为水平组合 $A_i B_j$ 的交互作用。它表示总效应去除分效应 α_i 及 β_j 后的剩余部分。此时有

$$\mu_{ij} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij}$$

交互作用满足

$$\sum_{j=1}^s (\alpha\beta)_{ij} = 0, \quad \sum_{i=1}^r (\alpha\beta)_{ij} = 0.$$
$$i = 1, 2, \dots, r, \quad j = 1, 2, \dots, s.$$

通常将因素 A 与 B 的交互作用看成一种因素, 记为 $A \times B$.

一 无交互作用的两因素方差分析

每种水平组合只进行一次试验，试验数据如下表

因素A \ 因素B	B_1, B_2, \dots, B_s	$\bar{X}_{i\cdot}$
A_1	$X_{11}, X_{12}, \dots, X_{1s}$	$\bar{X}_{1\cdot}$
A_2	$X_{21}, X_{22}, \dots, X_{2s}$	$\bar{X}_{2\cdot}$
\dots	$\dots\dots\dots$	\dots
A_r	$X_{r1}, X_{r2}, \dots, X_{rs}$	$\bar{X}_{r\cdot}$
$\bar{X}_{\cdot j}$	$\bar{X}_{\cdot 1}, \bar{X}_{\cdot 2}, \dots, \bar{X}_{\cdot s}$	\bar{X}

(一) 无交互作用的两因素方差分析模型

$$\left\{ \begin{array}{l} X_{ij} = \mu + \alpha_i + \beta_j + \varepsilon_{ij} \\ \varepsilon_{ij} \stackrel{i.i.d}{\sim} N(0, \sigma^2), \quad \begin{array}{l} i = 1, 2, \dots, r. \\ j = 1, 2, \dots, s \end{array} \\ \sum_{i=1}^r \alpha_i = \sum_{j=1}^s \beta_j = 0, \\ \mu, \alpha_i, \beta_j, \sigma^2 \text{ 未知.} \end{array} \right.$$

此模型的假设检验有两个

$$\left\{ \begin{array}{ll} H_{0A} : \alpha_1 = \dots = \alpha_r = 0, & H_{1A} : \alpha_1, \dots, \alpha_r \text{ 不全为零.} \\ H_{0B} : \beta_1 = \dots = \beta_s = 0, & H_{1B} : \beta_1, \dots, \beta_s \text{ 不全为零.} \end{array} \right.$$

数据 X_{ij} 的差异

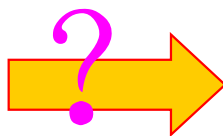
原因

$\boxed{\text{因素 A}} \oplus \boxed{\text{因素 B}} \oplus \boxed{\text{其他因素}}$

大小

$$S_T = \sum_{i=1}^r \sum_{j=1}^s (X_{ij} - \bar{X})^2$$

S_T 的分解



与因素 A
有关的项

\oplus

与因素 B
有关的项

\oplus

与其它因素
有关的项

比较大小

比较大小

(二) 平方和分解公式

$$\left\{ \begin{aligned} S_T &= \sum_{i=1}^r \sum_{j=1}^s (X_{ij} - \bar{X})^2 \\ &= \sum_{i=1}^r \sum_{j=1}^s \left[(X_{ij} - \bar{X}_{i\cdot} - \bar{X}_{\cdot j} + \bar{X}) + (\bar{X}_{i\cdot} - \bar{X}) + (\bar{X}_{\cdot j} - \bar{X}) \right]^2 \\ &= S_e + S_A + S_B \\ S_e &= \sum_{i=1}^r \sum_{j=1}^s (X_{ij} - \bar{X}_{i\cdot} - \bar{X}_{\cdot j} + \bar{X})^2 \\ S_A &= s \sum_{i=1}^r (\bar{X}_{i\cdot} - \bar{X})^2, \quad S_B = r \sum_{j=1}^s (\bar{X}_{\cdot j} - \bar{X})^2 \end{aligned} \right.$$

根据 $X_{ij} \sim N(\mu + \alpha_i + \beta_j, \sigma^2)$ 可得

$$\bar{X}_{i.} \sim N(\mu + \alpha_i, \frac{1}{s} \sigma^2), \quad \bar{X}_{.j} \sim N(\mu + \beta_j, \frac{1}{r} \sigma^2),$$

$$\bar{X} \sim N(\mu, \frac{\sigma^2}{rs}).$$

因而

$$X_{ij} - \bar{X}_{i.} - \bar{X}_{.j} + \bar{X} \sim N(0, \frac{(r-1)(s-1)}{rs} \sigma^2)$$

$$\Rightarrow E(S_e) = (r-1)(s-1)\sigma^2,$$

$$\bar{X}_{i.} - \bar{X} \sim N(\alpha_i, \frac{r-1}{rs} \sigma^2)$$

$$\Rightarrow E(S_A) = (r-1)\sigma^2 + s \sum_{i=1}^r \alpha_i^2,$$

$$\bar{X}_{\cdot j} - \bar{X} \sim N(\beta_j, \frac{s-1}{rs} \sigma^2),$$

$$\Rightarrow E(S_B) = (s-1)\sigma^2 + r \sum_{j=1}^s \beta_j^2.$$

根据上述特点, 两个假设检验的否定区域分别为

$$W_A = \left\{ F_A = \frac{S_A / r - 1}{S_e / (r-1)(s-1)} > C_1 \right\}$$

$$W_B = \left\{ F_B = \frac{S_B / s - 1}{S_e / (r-1)(s-1)} > C_2 \right\}$$

(三) 检验统计量与否定域

可以证明 $S_e / \sigma^2 \sim \chi^2((r-1)(s-1))$,

$$S_A/\sigma^2 \stackrel{H_{0A}}{\sim} \chi^2(r-1), \quad S_B/\sigma^2 \stackrel{H_{0B}}{\sim} \chi^2(s-1).$$

H_{0A}, H_{0B} 的检验统计量分别为

$$F_A = \frac{S_A/r-1}{S_e/(r-1)(s-1)} \stackrel{H_{0A}}{\sim} F(r-1, (r-1)(s-1))$$

$$F_B = \frac{S_B/s-1}{S_e/(r-1)(s-1)} \stackrel{H_{0B}}{\sim} F(s-1, (r-1)(s-1))$$

给定显著性水平 α 时, H_{0A}, H_{0B} 的否定域为

$$W_A = \left\{ F_A > F_\alpha(r-1, (r-1)(s-1)) \right\},$$

$$W_B = \left\{ F_B > F_\alpha(s-1, (r-1)(s-1)) \right\}.$$

（四）例子

例2 设有4名工人分别操作机床甲、乙、丙各一天，生产同种产品，其日产量统计如下表(单位：件). 问工人的不同和机床的不同对日产量有无显著性影响？($\alpha=0.05$ ，假定四名工人对这三台机床的熟悉情况是一样的).

工人 机床	张某	李某	王某	赵某
甲	53	47	57	45
乙	56	50	63	52
丙	45	47	54	42

解：把工人看成因素A，它有四个水平，把机床看成因素B，它有三个水平，由题意(假定四名工人对这三台机床的熟悉情况是一样的)知：因素A与因素B无交互作用. 根据公式计算 S_T , S_e , S_A , S_B 的值得到如下的方差分析表.

方差来源	平方和	自由度	均方和	F值	临界值	显著性
因素A	239.59	3	79.86	17.03	$F_{0.01}(3,6) = 9.78$	**
因素B	137.17	2	68.59	14.62	$F_{0.01}(2,6) = 10.9$	**
误差e	28.16	6	4.69			
总和	404.92	11				

结果表明:工人的不同和机床的不同对日产量有非常显著的影响,王某的产量和乙机床的产量比较高.

二 有交互作用的两因素方差分析模型

当存在交互作用时 $X_{ij} \sim N(\mu + \alpha_i + \beta_j + (\alpha\beta)_{ij}, \sigma^2)$

$$X_{ij} - \bar{X} \sim N(\alpha_i + \beta_j + (\alpha\beta)_{ij}, \frac{rs - 1}{rs} \sigma^2),$$

$$\bar{X}_{i.} - \bar{X} \sim N(\alpha_i, \frac{r - 1}{rs} \sigma^2), \quad \bar{X}_{.j} - \bar{X} \sim N(\beta_j, \frac{s - 1}{rs} \sigma^2),$$

$$X_{ij} - \bar{X}_{i.} - \bar{X}_{.j} + \bar{X} \sim N((\alpha\beta)_{ij}, \frac{(r - 1)(s - 1)}{rs} \sigma^2)$$

$$\Rightarrow E(S_e) = (r-1)(s-1)\sigma^2 + \sum_{i=1}^r \sum_{j=1}^s (\alpha\beta)_{ij}^2$$

为了得到随机误差项 S_e ，需要进行重复试验，假定每种水平组合试验 $t(t > 1)$ 次。试验数据如下表

因素A \ 因素B	$B_1 \cdots \cdots B_j \cdots \cdots B_s$	$\bar{X}_{i\cdot}$
A_1	$\bar{X}_{1\cdot}$
...
A_i	... $X_{ij1}, X_{ij2}, \cdots, X_{ijt}$...	$\bar{X}_{i\cdot}$
...
A_r	$\bar{X}_{r\cdot}$
$\bar{X}_{\cdot j}$	$\bar{X}_{\cdot 1} \cdots \bar{X}_{\cdot j} \cdots \bar{X}_{\cdot s}$	\bar{X}

$$\begin{aligned}
 \text{记 } \bar{X} &= \frac{1}{rst} \sum_{i=1}^r \sum_{j=1}^s \sum_{k=1}^t X_{ijk} & \bar{X}_{ij\cdot} &= \frac{1}{t} \sum_{k=1}^t X_{ijk} \\
 \bar{X}_{i..} &= \frac{1}{st} \sum_{j=1}^s \sum_{k=1}^t X_{ijk} & \bar{X}_{\cdot j\cdot} &= \frac{1}{rt} \sum_{i=1}^r \sum_{k=1}^t X_{ijk}
 \end{aligned}$$

(一) 方差分析模型

$$\left\{ \begin{aligned}
 &X_{ijk} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \varepsilon_{ijk}, \quad k = 1, 2, \dots, t. \\
 &\varepsilon_{ijk} \stackrel{i.i.d}{\sim} N(0, \sigma^2), \quad \sum_{i=1}^r \alpha_i = \sum_{j=1}^s \beta_j = 0, \\
 &\sum_{i=1}^r (\alpha\beta)_{ij} = \sum_{j=1}^s (\alpha\beta)_{ij} = 0, \quad \begin{aligned} &i = 1, 2, \dots, r, \\ &j = 1, 2, \dots, s. \end{aligned} \\
 &\mu, \alpha_i, \beta_j, (\alpha\beta)_{ij}, \sigma^2 \text{ 未知.}
 \end{aligned} \right.$$

此模型的假设检验有三个

$$\left\{ \begin{array}{ll} H_{0A} : \alpha_1 = \cdots = \alpha_r = 0, & H_{1A} : \alpha_1, \cdots, \alpha_r \text{不全为零.} \\ H_{0B} : \beta_1 = \cdots = \beta_s = 0, & H_{1B} : \beta_1, \cdots, \beta_s \text{不全为零.} \\ H_{0A \times B} : \text{所有 } (\alpha\beta)_{ij} = 0, & H_{1A \times B} : (\alpha\beta)_{ij} \text{不全为零.} \end{array} \right.$$

数据 X_{ijk} 的差异

原因

$$\boxed{\text{因素A}} \oplus \boxed{\text{因素B}} \oplus \boxed{\text{交互作用}} \oplus \boxed{\text{其他因素}}$$

大小

$$S_T = \sum_{i=1}^r \sum_{j=1}^s \sum_{k=1}^t (X_{ijk} - \bar{X})^2 \quad ? \text{分解}$$

(二) 平方和分解公式

$$\left\{ \begin{array}{l} S_T = \sum_{i=1}^r \sum_{j=1}^s \sum_{k=1}^t (X_{ijk} - \bar{X})^2 = S_e + S_A + S_B + S_{A \times B} \\ S_e = \sum_{i=1}^r \sum_{j=1}^s \sum_{k=1}^t (X_{ijk} - \bar{X}_{ij\cdot})^2 \\ S_A = st \sum_{i=1}^r (\bar{X}_{i..} - \bar{X})^2 \quad S_B = rt \sum_{j=1}^s (\bar{X}_{.j\cdot} - \bar{X})^2 \\ S_{A \times B} = t \sum_{i=1}^r \sum_{j=1}^s (\bar{X}_{ij\cdot} - \bar{X}_{i..} - \bar{X}_{.j\cdot} + \bar{X})^2 \end{array} \right.$$

$$S_T = \sum_{i=1}^r \sum_{j=1}^s \sum_{k=1}^t (X_{ijk} - \bar{X})^2 = \sum_{i=1}^r \sum_{j=1}^s \sum_{k=1}^t X_{ijk}^2 - \frac{1}{rst} \left(\sum_{i=1}^r \sum_{j=1}^s X_{ijk} \right)^2$$

(三) 检验统计量与否定域

$$\text{由于 } S_e/\sigma^2 \sim \chi^2(rs(t-1)), \quad S_A/\sigma^2 \stackrel{H_{0A}}{\sim} \chi^2(r-1), \\ S_B/\sigma^2 \stackrel{H_{0B}}{\sim} \chi^2(s-1). \quad S_{A \times B}/\sigma^2 \stackrel{H_{0A \times B}}{\sim} \chi^2((r-1)(s-1)).$$

因此 $H_{0A}, H_{0B}, H_{0A \times B}$ 的检验统计量分别为

$$F_A = \frac{S_A/r-1}{S_e/rs(t-1)} \stackrel{H_{0A}}{\sim} F(r-1, rs(t-1)),$$

$$F_B = \frac{S_B/s-1}{S_e/rs(t-1)} \stackrel{H_{0B}}{\sim} F(s-1, rs(t-1)),$$

$$F_{A \times B} = \frac{S_{A \times B}/(r-1)(s-1)}{S_e/rs(t-1)} \stackrel{H_{0A \times B}}{\sim} F((r-1)(s-1), rs(t-1)).$$

给定显著性水平 α 时, $H_{0A}, H_{0B}, H_{0A \times B}$ 的否定域为

$$W_A = \{F_A > F_\alpha(r-1, rs(t-1))\}$$

$$W_B = \{F_B > F_\alpha(s-1, rs(t-1))\}$$

$$W_{A \times B} = \{F_{A \times B} > F_\alpha((r-1)(s-1), rs(t-1))\}$$

若 $F > F_\alpha$, 则认为因素取不同水平(或交互作用)对指标影响显著.

$F > F_{0.01}$, 认为影响是高度显著的, 用**表示;

$F_{0.01} < F \leq F_{0.05}$, 认为影响是显著的, 用*表示;

$F_{0.05} < F \leq F_{0.1}$, 认为有一定显著, 用(*)表示;

$F < F_{0.1}$, 认为影响不显著, 无表示.

(四) 例子

例2 为了比较三种松树在四个不同的地区的生长情况有无差别，在每个地区对每种松树随机地选取**5株**，测得它们的胸径，得到如下数据.(单位：**cm**)

	地 区			
松树品种	1	2	3	4
1	23 15 26 13 21	25 20 21 16 18	21 17 16 24 27	14 17 19 20 24
2	28 22 25 19 26	30 26 26 20 28	19 24 19 25 29	17 21 18 26 23
3	18 10 12 22 13	15 21 22 14 12	23 25 19 13 22	18 12 23 22 19

解：把树种看成因素A，它有3个水平，把地区看成因素B，它有4个水平，两种因素可能存在交互作用．根据公式计算 S_T ， S_e ， S_A ， S_B ， S_{A*B} 的值得到如下的方差分析表．

方差来源	平方和	自由度	均方和	F值	临界值	显著性
因素A	355.6	2	177.8	9.68	F_{0.01} (2,48) =5.08	* *
因素B	49.65	3	16.55	0.9	F_{0.01} (3,48) =4.22	
交互作用 A*B	106.4	6	17.73	0.97	F_{0.01} (6,48) =3.20	
误差e	882	48	18.38			
总和	1393.7	11				

由方差分析表知：树种对松树的影响程度是高度显著的，而地区和交互作用不显著。

单元均值	B1	B2	B3	B4	A水平均值
A1	19.6	20.0	21.0	18.8	19.85
A2	24.0	26.0	23.2	21.0	23.55
A3	15.0	16.8	20.4	18.4	17.65
B水平均值	19.53	20.93	21.53	19.40	20.35

进一步考察因素A不同水平的均值, 可知树种2的均值最大，因此树种2的生长优于其他的树种.