第六章 方差分析基础

揚州大學
YANGZHOU UNIVERSITY

6.1方差分析的意义

从第5章可知，至多两个平均数的假设测验根据具体情况可选用 u 测验或小样本 t 测验。但是，多个平均数（$k \geq 3$）比较在农业生产和科学研究中比较普遍，需要进行多个平均数间的假设测验。这时，不宜采用 u 或 t 测验，原因为：

① 程序繁琐。

k 个处理平均数间的两两比较，需要进行 $C_k^2 = k(k-1)/2$ 次 u 或 t 检验

例如 $k = 10$，就要进行 1/2×10×9=45次测验，假设测验过程繁琐

6.1方差分析的意义

② 随着试验的增大，会大大增加犯 α 错误的概率。

若每次检验犯第 I 类错误的概率为 α，则 C_k^2 次检验犯第I类错误概率为 $\alpha' = 1-(1-\alpha)^{C_k^2}$

当 $k = 4$ 和 $\alpha = 0.05$时，则 $\alpha' = 1-(1-0.05)^6 = 0.2649$ ，远大于0.05；

③ 误差估计精确度受到损失。

只利用两样本信息估计误差方差，误差自由度$df_e = 2\ (n\text{-}1)$；而利用k个样本联合估计误差方差时，误差自由度$df_e = k\ (n\text{-}1)$。显然，随着误差自由度增大，提高了误差方差估计精度和假设检验的灵敏性。

6.1方差分析的意义

为克服上述问题，多个平均数的假设检验应当采用英国统计学家 **R. A. Fisher** 提出的方差分析方法 **(analysis of variance，ANOVA)**。

基本原理：将所有观察值作为一个整体，把观察值总变异的平方和和自由度分解为不同变异来源的平方和和自由度，进而获得不同变异来源的总体方差估值，然后计算这些估值的 F 值，就能测验假设 H_0: $\mu_1 = \mu_2 = \ldots = \mu_k$(各样本的总体平均数相等)。

方差分析在科研工作中有极重要的用途。特别在多因素试验中，它可以帮助我们发现起主导作用的变异来源，从而抓住主要矛盾和关键措施。本章主要讨论基础的方差分析方法；更复杂的情形将在第七章与试验设计一起讨论。

第六章 方差分析基础

6.2 单向分组资料的方差分析

设有 k 个独立的样本，每个样本具有 n 个观察值。该资料称为具有 kn 个观察值的单向分组资料，其数据模式如下表：

表6.1 kn 个观察值的单向分组资料的数据模式

组别	观察值 (y_{ij}， i =1，2，…，k； j = 1,2…，n)						总和	平均
1	y_{11}	y_{12}	…	y_{1j}	…	y_{1n}	T_1	$\bar{y}_1$
2	y_{21}	y_{22}	…	y_{2j}	…	y_{2n}	T_2	$\bar{y}_2$
			…		…			
i	y_{i1}	y_{i2}	…	y_{ij}	…	y_{in}	T_i	$\bar{y}_i$
			…		…			
k	y_{k1}	y_{k2}	…	y_{kj}	…	y_{kn}	T_k	$\bar{y}_k$
							T	$\bar{y}$

$T=\sum Y_{ij}=\sum T_i$，即全部观察值的总和数； $\bar{y}=\sum Y_{ij}/kn$，即全部观察值的平均数。

6.2 单向分组资料的方差分析

平方和与自由度的分解

总变异可分解为组间（处理间）变异和组内（试验误差）变异。

总变异的平方和 SS_T 是各观察值 y_{ij} 与 $\bar{y}$ 的离差平方之和：

$$SS_T = \sum_{1}^{nk}(y_{ij} - \bar{y})^2 = \sum_{1}^{nk} y_{ij}^2 - C$$

其中的C称为矫正数： $C = \frac{T^2}{nk}$

其自由度为： $df_T = nk\text{-}1$

6.2 单向分组资料的方差分析

各样本内（处理内）平方和 SS_i 是各样本的观察值与其平均数 $\bar{y}_i$ 的离差平方之和。

对于样本 1 是： $SS_1=\sum_{1}^{n}\left(y_{1j}-\bar{y}_1\right)^2$ $df_1=(n-1)$

对于样本 2 是： $SS_2=\sum_{1}^{n}\left(y_{2j}-\bar{y}_2\right)^2$ $df_2=(n-1)$

对于样本 k 是： $SS_k=\sum_{1}^{n}\left(y_{kj}-\bar{y}_k\right)^2$ $df_k=(n-1)$

SS_1，SS_2，… SS_k 都属随机误差平方和的估值，因而可将其合并以提供随机误差平方和（或样本内平方和）的更精确估值：

$$SS_e=SS_1+SS_2+\cdots+SS_k=\sum_{1}^{k}\sum_{1}^{n}\left(y_{ij}-\bar{y}_i\right)^2$$

SS_e的相应自由度则为： $df_e=df_1+df_2+\cdots+df_k=k(n-1)$

6.2 单向分组资料的方差分析

样本（处理）平均数间的平方和 SS_t 是各样本的平均数 $\bar{y}_i$ 对 $\bar{y}$ 的离差平方之和：

$$SS_t = \sum_1^k (\bar{y}_i - \bar{y})^2$$

其**自由度**为： $df_t = (k-1)$

SS_t/df_t 系样本平均数的方差，是 σ^2/n 的估值；而 SS_e/df_e 则是 σ^2 的估值。

为了进行正确的 F 测验，必须使它们都是估计着同一参数 σ^2 。因而，样本间的平方和应为：

$$SS_t = n\sum_1^k (\bar{y}_i - \bar{y})^2 = \frac{\sum_1^k T_i^2}{n} - C$$

6.2 单向分组资料的方差分析

至此，我们完成了表6-1数据资料的平方和和自由度的分解，并得到恒等式：

总变异的 = 样本间的 + 样本内的

$$\left.\begin{aligned}&\sum_{1}^{kn}(Y_{ij}-\bar{y})^2=n\sum_{1}^{k}(\bar{y}_i-\bar{y})^2+\sum_{1}^{k}\sum_{1}^{n}\left(Y_{ij}-\bar{y}_i\right)^2\\&(kn-1)=(k-1)+k(n-1)\end{aligned}\right\}$$

进而得到样本间均方 MS_t 和样本内均方 MS_e 为：

$$\left.\begin{aligned}MS_t&=n\sum_{1}^{k}(\bar{y}_i-\bar{y})^2\Big/(k-1)\\MS_e&=\sum_{1}^{k}\sum_{1}^{n}\left(Y_{ij}-\bar{y}_i\right)^2\Big/k(n-1)\end{aligned}\right\}$$

6.2 单向分组资料的方差分析

如果样本平均数间的变异并不显著大于样本内（随机误差的变异），则 $F = MS_t / MS_e$ 将不会给出显著的值（这时F 的期望值是 ）；反之，则 F 将给出显著的值。所以，由

$$F = MS_t / MS_e$$

可测验 H_0：$\mu_1 = \mu_2 = \ldots = \mu_k$ (简写作 H_0：$\mu_i = \mu$)，对 H_A：μ_1， μ_2， … μ_k 不相等。

6.2 单向分组资料的方差分析

表6.2 单向分组资料的方差分析表

变异来源	*df*	*SS*	*MS*	*F*
处理间	$k-1$	$n\sum_{1}^{k}(\bar{y}_i-\bar{y})^2$	MS_t	MS_t/MS_e
处理内	$k(n-1)$	$\sum_{1}^{k}\sum_{1}^{n}(Y_{ij}-\bar{y}_i)^2$	MS_e	
总变异	$kn-1$	$\sum_{1}^{kn}(Y_{ij}-\bar{y})^2$		

6.2 单向分组资料的方差分析

例6-1 利用4种培养基A、B、C和D进行组织培养，每处理各得4个苗高观察值（mm），其结果如表6-3。试进行平方和分解。

表6.3 不同培养基的苗高（单位：mm）

培养基	观察值				总和	平均数
A	18	21	20	13	72	18
B	20	24	26	22	92	23
C	10	15	17	14	56	14
D	28	27	29	32	116	29
					336	21

6.2 单向分组资料的方差分析

扬州大学 YANGZHOU UNIVERSITY

矫正数 $C=\dfrac{T^2}{nk}=\dfrac{336^2}{4\times 4}$=7056

总平方和 $SS_T=\sum_{i=1}^{4}\sum_{j=1}^{4}y_{ij}^2-C=18^2+21^2+\dots+32^2-C=602$

处理平方和 $SS_t=n\sum_1^k(\bar{y}_i-\bar{y})^2=\sum T_i^2/n-C$

$$=(72^2+92^2+56^2+116^2)/4-C=504$$

误差平方和 $SS_e=\sum_1^k\sum_1^n(y_{ij}-\bar{y}_i)^2=\sum_1^{nk}y_{ij}^2-\sum_1^k T_i^2/n$

$$=SS_T-SS_t=602-504=98$$

6.2 单向分组资料的方差分析

例6-2 作一水稻施肥的盆栽试验，设5个处理，A和B系分别施用两种不同工艺流程的氨水，C施碳酸氢铵，D施尿素，E不施氮肥。每处理4盆(施肥处理的施肥量每盆皆为折合纯氮1.2克)，共5×4=20盆，随机放置于同一网室中，其稻谷产量(克/盆)列于表6.4，试测验各处理平均数的差异显著性。

表6.4 水稻施肥盆栽试验的产量结果

肥料种类	观察值				平均数	总和
A	24	30	28	26	27	108
B	27	24	21	26	24.5	98
C	31	28	25	30	28.5	114
D	32	33	33	28	31.5	126
E	21	22	16	21	20	80
					26.3	526

6.2 单向分组资料的方差分析

自由度和平方和的分解

总变异自由度 $df_T = nk－1 = 5×4－1 = 19$

处理间自由度 $df_t = k－1 = 5－1 = 4$

误差(处理内)自由度 $df_e = k(n－1) = 5×(4－1) = 15$

矫正数 $C = T^2/nk = 526^2/(5\times 4) = 13833.8$

$$SS_T = \sum y^2 - C = 24^2 + 30^2 + \cdots + 21^2 - C = 402.2$$

$$SS_t = \sum T_i^2/n - C = (108^2 + 98^2 + \cdots + 80^2)/4 - C = 301.2$$

$$SS_e = 402.2 - 301.2 = 101.0$$

6.2 单向分组资料的方差分析

表6.5 表6-4资料的方差分析

变异来源	df	SS	MS	F	$F_{0.05}$	$F_{0.01}$
处 理 间	4	301.2	75.30	11.19**	3.06	4.89
处理内 (试验误差)	15	101.0	6.73			
总 变 异	19	402.2				

为了测验 H_0: $\mu_A = \mu_B = \ldots = \mu_E$ (即5个处理的总体平均数相等)，计算处理间均方对误差均方的比率，得到F = 75.3/6.73 = 11.19。查附表v_1=4, v_2=15, $F_{0.01}$=4.89。 F = 11.19 > $F_{0.01}$， 故否定 **H_0 推断这个试验的处理平均数间是有极显著差异的。**

第六章 方差分析基础

6.3 多重比较

F 测验是一种整体性质的测验。一个显著的F 值只表明试验中的各个处理平均数间存在显著差异，而不是任何两个平均数间都有显著差异，例如，表6.4资料通过表6.5的方差分析，否定了H_0: $\mu_A = \mu_B = \ldots = \mu_E$ ，但这并不意味着接受 $\mu_A \neq \mu_B \neq \ldots \neq \mu_E$ ，因为 H_A 是 μ_A，μ_B，… , μ_E不相等。

为了明确哪些平均数间的差数为显著，哪些为不显著，需进一步对各个平均数作相互比较，即多重比较（multiple comparisons）。

多重比较的方法甚多，这里主要介绍3种：

1、最小显著差数法 (简称LSD法)

2、Duncan 氏新复极差测验法 (简称SSR法)

3、Tukey 氏固定极差测验法 (简称HSD法)

6.3 多重比较

1、最小显著差数法 (简称LSD法)

LSD法实质上是第五章的 t 测验。其程序为：在处理间的F 测验为显著的前提下，计算出显著水平为α的最小显著差数LSD_{α}；任何两个处理平均数的差数 $\bar{y}_i - \bar{y}_j$，如其绝对值> LSD_{α}，即为在α水平上显著；反之，则为在α水平上不显著，若处理间的F不显著则接受H_0: $\mu_i = \mu$，即推断各处理的平均数间均无显著差异，不再进行多重比较。

已知： $$t = \frac{\bar{y}_i - \bar{y}_j}{s_{\bar{y}_i - \bar{y}_j}} (i, j = 1, 2, \cdots, k, i \neq j)$$

若 $|t| \geq t_{\alpha}$， $\bar{y}_i - \bar{y}_j$ 即为在 α 水平上显著。

因此，最小显著差数为： $LSD_{\alpha} = t_{\alpha} s_{\bar{y}_i - \bar{y}_j}$

当两样本的容量 n 相等时，$s_{\bar{y}_i - \bar{y}_j} = \sqrt{2MS_e / n}$

6.3 多重比较

例6-3 试以LSD法测验例6-2各处理平均数间的差异显著性。

由表6.5已知 $MS_e = 6.73$，$df_e=15$，故

$$s_{\bar{y}_i-\bar{y}_j} = \sqrt{\frac{2 \times 6.73}{4}} = 1.834$$

查表得，$\nu = df_e = 15$ $t_{0.05}=2.131$ $t_{0.01}=2.947$

$$LSD_{0.05} = t_{0.05} \cdot s_{\bar{y}_i-\bar{y}_j} = 2.131 \times 1.834 = 3.90$$

$$LSD_{0.01} = t_{0.01} \cdot s_{\bar{y}_i-\bar{y}_j} = 2.947 \times 1.834 = 5.40$$

这是说，表6-4的任何两个处理平均数，其差数的绝对值如 >3.90克为 $\alpha = 0.05$ 水平上显著，如 >5.40克为 $\alpha = 0.01$ 水平上显著。

表6.6 例6-2资料的差异显著性

肥料种类	平均数	差异显著性	
		0.05	**0.01**
D	31.5	a	A
C	28.5	ab	AB
A	27.0	bc	AB
B	24.5	c	BC
E	20.0	d	C

6.3 多重比较

字母标记法：

① 平均数按大小排列

② 最高之处先标a

③ 无差异同字母

④ 有了差异换字母

⑤ 换了字母要回头

⑥ 相同字母取高处

$LSD_{0.05} = 3.90$

$LSD_{0.01} = 5.40$

肥料种类	平均数	差异显著性	
		0.05	**0.01**
D	31.5	a	A
C	28.5	ab	AB
A	27.0	bc	AB
B	24.5	c	BC
E	20.0	d	C

在表示多重比较结果时，可以采用字母标记法，给没有显著差异的处理以相同字母，给有显著差异的处理以不同字母。这样可以紧凑的篇幅而使各处理间的差异显著性一目了然。

6.3 多重比较

2、Duncan 氏新复极差测验法 (简称SSR法)

这是D.B.Duncan（1995）提出的一种多重比较方法。这种测验法以比较错误率为标准，又叫最短显著极差（shortest significant ranges，简记作*SSR*）法。

其特点是：依平均数秩次距的不同而采用一系列不同的显著值。这些显著值叫做多重极差（multiple range），记作*LSR*或*R*。

$$R_\alpha = SSR_\alpha \cdot s_{\bar{y}}$$

$s_{\bar{y}}$ 为平均数的标准误: $s_{\bar{y}} = \sqrt{\frac{MS_e}{n}}$

D
C
A
B
E

SSR_α 是自由度为 df_e、秩次距为k 和显著性水平为α条件下的临界*SSR*值，可查附表6得到。

秩次距是所有比较平均数按大到小的顺序排列所计算出的两极差范围内所包含的平均数个数。

例6-4 试以*SSR*法测验例6-2各处理平均数间的差异显著性。

由表6.5已知 $MS_e = 6.73$，n =4 故

$$s_{\bar{y}} = \sqrt{\frac{6.73}{4}} = 1.297$$

$v = df_e$=15，在附表6分别查出 k = 2, 3, 4, 5时的$SSR_{0.05}$ 和$SSR_{0.01}$

表6.7 例6-2资料*SSR*值的计算

秩次距	$SSR_{0.05}$	$R_{0.05}$	$SSR_{0.01}$	$R_{0.01}$
2	3.01	3.9	4.17	5.41
3	3.16	4.1	4.37	5.64
4	3.25	4.22	4.5	5.78
5	3.31	4.29	4.58	5.9

D
C
A
B
E

6.3 多重比较

表6.8 例6-2资料的差异显著性（*SSR*测验）

肥料种类	平均数	差异显著性	
		0.05	0.01
D	31.5	a	A
C	28.5	ab	AB
A	27.0	b	AB
B	24.5	b	BC
E	20.0	c	C

6.3 多重比较

3、Tukey 氏固定极差测验法 (简称HSD法)

将所有两两平均数间的差数与同一个固定极差D_α作比较。

$$D_\alpha = q_{\alpha,k,df_e} \cdot s_{\bar{y}}$$

$s_{\bar{y}}$ 为平均数的标准误: $s_{\bar{y}} = \sqrt{\frac{MS_e}{n}}$

q_α的值随k和df_e而异，k为最大的秩次距，可通过查附表5获得。

与SSR法的区别在于，LSR_α随秩次距k的不同而异，而D_α是采用了一个固定的秩次距。

6.3 多重比较

多重比较方法的选择

一个试验资料，采用哪一种多重比较方法，主要应根据否定一个正确的H_0和接受一个不正确的H_0的相对重要性决定。

- 如果否定正确的H_0是事关重大或后果严重的，应用Tukey测验法。
- 如果接受不正确的H_0是事关重大或后果严重的，则宜应用*LSD*测验和*SSR*测验。

由于试验工作者通常都寄希望于否定H_0所以*LSD*测验和*SSR*测验得到较为广泛的应用。

生物试验中，由于试验误差较大，常采用*LSD*法。

第六章 方差分析基础

- 6.1 方差分析的意义
- 6.2 单向分组资料的方差分析
- 6.3 多重比较
- **6.4 两向分组资料的方差分析**
- 6.5 系统分组资料的方差分析
- 6.6 变量转换

6.4 两向分组资料的方差分析

两因素试验中若因素A的每个水平与因素B的每个水平均衡相遇(或称正交)，则所得试验数据按两个因素交叉分组称为两向分组资料。

6.4.1 无重复观察值两向分组资料的方差分析

6.4.2 有重复观察值两向分组资料的方差分析

6.4.1 无重复观察值两向分组资料的方差分析

设有*A*、*B*两个试验因素，*A*因素有*a*个水平，*B*因素有*b*个水平，则有*ab*个处理组合，每个组合有1个观察值，则该试验共有*ab*个观测值。试验资料模式如表6-9所示。

表6.9 无重复观察值的两因素双向分组资料数据模式

（i=1，2，…，a；j=1，2，…，b）

因素 A	因素 B				总和 $T_{i.}$	平均 $\bar{y}_{i.}$
	B_1	B_2	…	B_b		
A_1	y_{11}	y_{12}	…	y_{1b}	$T_{1.}$	$\bar{y}_{1.}$
A_2	y_{21}	y_{22}	…	y_{2b}	$T_{2.}$	$\bar{y}_{2.}$
⋮	⋮	⋮	⋮	⋮	⋮	⋮
A_a	y_{a1}	y_{a2}	…	y_{ab}	$T_{a.}$	$\bar{y}_{a.}$
总和 $T_{.j}$	$T_{.1}$	$T_{.2}$	…	$T_{.b}$	T	
平均 $\bar{y}_{.j}$	$\bar{y}_{.1}$	$\bar{y}_{.2}$	…	$\bar{y}_{.b}$		$\bar{y}_{..}$

6.4.1 无重复观察值两向分组资料的方差分析

表6.10 表6.9类型资料自由度和平方和的分解及方差分析

变异来源	DF	SS	MS	F
A 因素	$df_a = a-1$	$SS_A = b\sum(\bar{y}_{i.} - \bar{y}_{..}) = \dfrac{\sum T_{i.}^2}{b} - C$	$MS_A = \dfrac{SS_A}{df_A}$	$\dfrac{MS_A}{MS_e}$
B 因素	$df_b = b-1$	$SS_B = a\sum(\bar{y}_{.j} - \bar{y}_{..}) = \dfrac{\sum T_{.j}^2}{a} - C$	$MS_B = \dfrac{SS_B}{df_B}$	$\dfrac{MS_B}{MS_e}$
误差	$df_e = (a-1)(b-1)$	$SS_e = SS_T - SS_A - SS_B$	$MS_e = \dfrac{SS_e}{df_e}$	
总变异	$df_T = ab-1$	$SS_T = \sum(y_{ij} - \bar{y}_{..}) = \sum y^2 - C$		

6.4.1 无重复观察值两向分组资料的方差分析

【例6-5】在5种不同温度下研究柑橘黑色蒂腐病菌的生长和温度的关系，于接种后不同天数测量其生长速度，获得数据如下表（表6-11），试进行方差分析。

温度	接种后天数				总和 $T_{i.}$	平均 $\bar{y}_{i.}$
	B_1(1 天)	B_2(2 天)	B_3(3 天)	B_4(4 天)		
A_1（17.5）	0.3	1.3	2.6	3.5	7.7	1.93
A_2（21.0）	0.3	1.7	2.9	4.0	8.9	2.23
A_3（24.5）	0.9	3.0	6.6	7.5	18.0	4.50
A_4（27.5）	1.7	4.8	9.0	9.0	24.5	6.13
A_5（30.5）	1.2	2.7	5.2	7.4	16.5	4.13
总和 $T_{.j}$	4.4	13.5	26.3	31.4	T=75.6	
平均 $\bar{y}_{.j}$	0.88	2.70	5.26	6.28		$\bar{y}_{..}=3.78$

6.4.1 无重复观察值两向分组资料的方差分析

此例中 A 因素（温度）有 5 个水平，即 a=5，B 因素（接种后天数）有 4 个水平，即 b=4，共有 ab=5×4=20 个观测值。

（1）自由度和平方和的分解

总自由度 $df_T = ab-1 = 5\times 4-1 = 19$

A 因素自由度 $df_a = a-1 = 5-1 = 4$

B 因素自由度 $df_b = b-1 = 4-1 = 3$

误差自由度 $df_e = (a-1)(b-1) = (5-1)(4-1) = 12$

矫正数 $C = \dfrac{T^2}{\text{ab}} = \dfrac{75.6^2}{20} = 285.7680$

总平方和 $SS_T = \sum(y_{ij}-\bar{y}_{..}) = \sum y^2 - C = 0.3^2 + 1.3^2 + \cdots + 7.4^2 - 285.7680 = 150.4920$

A 因素平方和 $SS_A = b\sum(\bar{y}_{i.}-\bar{y}_{..}) = \dfrac{\sum T_{i.}^2}{b} - C = \dfrac{7.7^2+8.9^2+\cdots+16.5^2}{4} - 150.4920 = 47.9820$

B 因素平方和 $SS_B = a\sum(\bar{y}_{.j}-\bar{y}_{..}) = \dfrac{\sum T_{.j}^2}{a} - C = \dfrac{4.4^2+13.5^2+\cdots+31.4^2}{5} - 150.4920 = 90.0840$

误差平方和 $SS_e = SS_T - SS_A - SS_B = 150.4920 - 47.9820 - 90.0840 = 12.4260$

6.4.1 无重复观察值两向分组资料的方差分析

（2）计算方差（均方）

$$MS_A = \frac{SS_A}{df_A} = \frac{47.9820}{5} = 11.9955$$

$$MS_B = \frac{SS_B}{df_B} = \frac{90.0840}{4} = 30.0280$$

（3）列方差分析表，并进行 F 检验

表6.11 5 种不同温度与接种后天数柑橘黑色蒂腐病菌生长速度的方差分析表

变异来源	*DF*	*SS*	*MS*	*F*	$F_{0.05}$	$F_{0.01}$
A（温度）	4	47.9820	11.9955	11.58**	3.26	5.41
B（接种后天数）	3	90.0840	30.0280	29.00**	3.49	5.95
误差	12	12.4260	1.0355			
总变异	19	150.4920				

A因素（温度间）的$F > F_{0.01(4,\ 12)}=5.41$，$P < 0.01$，表明温度间差异极显著；$B$因素（接种后天数间）的$F > F_{0.01(3,\ 12)}=5.59$，$P < 0.01$，表明接种后天数间差异极显著。

6.4.1 无重复观察值两向分组资料的方差分析

（4）多重比较

①不同温度对柑橘黑色蒂腐病菌的生长速度的多重比较，采用 *SSR* 法。因为 *A* 因素（温度）每一水平的重复数为 *B* 因素的水平数 *b*，则 *A* 因素各水平的标准误为：

$$S_{\bar{y}} = \sqrt{\frac{MS_e}{b}} = \sqrt{\frac{1.0355}{4}} = 0.5088$$

当 df_e=12，a=2、3、4、5，查附表 6，得 α=0.05，α=0.01 的临界 *SSR* 值，用公式 $R_\alpha = S_{\bar{y}} \times SSR_\alpha$ 计算出最小显著极差时 *R*。*SSR* 值及 *R* 值列于表 6-12。

表6.12 多重比较时的R_α值计算

秩次距	2	3	4	5
$SSR_{0.05}$	3.08	3.23	3.33	3.36
$SSR_{0.01}$	4.32	4.55	4.68	4.76
$LSR_{0.05}$	1.57	1.64	1.69	1.71
$LSR_{0.01}$	2.20	2.32	2.38	2.42

6.4.1 无重复观察值两向分组资料的方差分析

表6.13 5种不同温度对柑橘黑色蒂腐病菌生长速度的显著性（*SSR*检验）

温度	菌落直径/ mm	差异显著性	
		5%	1%
A_4	6.13	a	A
A_3	4.50	b	AB
A_5	4.12	b	ABC
A_2	2.23	c	BC
A_1	1.93	c	C

温度间多重比较结果解释：温度A_4时微生物生物最快，极显著高于A_2、A_1，显著高于A_3、A_5；温度A_3极显著高于A_1，显著高于A_2；A_5显著高于A_2、A_1；A_3、A_5间差异不显著；A_2、A_1间差异不显著。

6.4.1 无重复观察值两向分组资料的方差分析

（4）多重比较

②接种后天数对柑橘黑色蒂腐病菌的生长速度的多重比较，采用 *SSR* 法。因为 *B* 因素（温度）每一水平的重复数为 *A* 因素的水平数 *a*，则 *B* 因素各水平的标准误为：

$$S_{\bar{y}}=\sqrt{\frac{MS_e}{a}}=\sqrt{\frac{1.0355}{5}}=0.4551$$

当 df_e=12，*a*=2、3、4、5，查附表 6，得 α=0.05，α=0.01 的临界 *SSR* 值，用公式 $R_\alpha=S_{\bar{y}}\times SSR_\alpha$ 计算出最小显著极差时 *LSR*。*SSR* 值及 *LSR* 值列于表 6-13。

表6.13 多重比较时的R_α值计算

秩次距	2	3	4	5
$SSR_{0.05}$	3.08	3.23	3.33	3.36
$SSR_{0.01}$	4.32	4.55	4.68	4.76
$LSR_{0.05}$	1.40	1.47	1.52	1.53
$LSR_{0.01}$	1.97	2.07	2.13	2.17

6.4.1 无重复观察值两向分组资料的方差分析

表6.14 接种后不同天数对柑橘黑色蒂腐病菌生长速度的显著性（*SSR*检验）

接种后天数	菌落直径/mm	差异显著性	
		5%	1%
B_4	6.28	a	A
B_3	5.26	a	A
B_2	2.70	b	B
B_1	0.88	c	B

温度间多重比较结果解释：接种后天数B_4微生物生长速度最快，极显著高于B_2、B_1；B_3极显著高于B_2、B_1；B_2显著高于B_1；B_4、B_3间差异不显著。

6.4.2 有重复观察值两向分组资料的方差分析

在进行两因素试验时，如果两因素存在互作，在进行方差分析时，需要将互作项和误差项的自由度和平方和分解开来，所以，进行两因素或多因素试验时应设置重复。

对于有重复观测值双向分组资料的方差分析，能研究因素的简单效应、主效和因素间的互作效应。

设有A、B两个试验因素，A因素有a个水平，B因素有b个水平，则有ab个处理组合，每个组合有n个观察值，则该试验共有abn个观测值。如果试验采用完全随机设计，则试验资料模式如表6-15所示。

6.4.2 有重复观察值两向分组资料的方差分析

表6.15 有重复观测值双向分组资料的数据模式

因素 A	因素 B				总和 $T_{i.}$	平均 $\bar{y}_{i.}$
	B_1	B_2	…	B_b		
A_1	y_{111} y_{112} ⋮ y_{11n}	y_{121} y_{122} ⋮ y_{12n}	… … ⋮ …	y_{1b1} y_{1b2} ⋮ y_{1bn}	$T_{1.}$	$\bar{y}_{1.}$
A_2	y_{211} y_{212} ⋮ y_{21n}	y_{221} y_{222} ⋮ y_{22n}	… … ⋮ …	y_{2b1} y_{2b2} ⋮ y_{2bn}	$T_{2.}$	$\bar{y}_{2.}$
⋮	⋮	⋮	⋮	⋮	⋮	⋮
A_a	y_{a11} y_{a12} ⋮ y_{a1n}	y_{a21} y_{a22} ⋮ y_{a2n}	… … ⋮ …	y_{ab1} y_{ab2} ⋮ y_{abn}	$T_{a.}$	$\bar{y}_{a.}$
总和 $T_{.j}$	$T_{.1}$	$T_{.2}$	…	$T_{.b}$	T	
平均 $\bar{y}_{.j}$	$\bar{y}_{.1}$	$\bar{y}_{.2}$	…	$\bar{y}_{.b}$		$\bar{y}_{..}$

***A B* 两向表**

因素 A	因素 B				总和
	B_1	B_2	…	B_b	$T_{i.}$
A_1	T_{11}	T_{12}	…	T_{1b}	$T_{1.}$
A_2	T_{21}	T_{22}	…	T_{2b}	$T_{2.}$
⋮	⋮	⋮	⋮	⋮	⋮
A_a	T_{a1}	T_{a2}	…	T_{ab}	$T_{a.}$
总和 $T_{.j}$	$T_{.1}$	$T_{.2}$	…	$T_{.b}$	T

总变异可分解为*A*因素水平间变异、*B*因素水平间变异、*AB*互作及试验误差。

6.4.2 有重复观察值两向分组资料的方差分析

表6.16 有重复观测值双向分组资料的平方和与自由度的分解表达式

变异来源	DF	SS	MS
处理组合	$df_t = ab-1$	$SS_t = \frac{\sum T_{ij.}^2}{n} - C$	$MS_t = \frac{SS_t}{df_t}$
A 因素	$df_a = a-1$	$SS_A = \frac{\sum T_{i..}^2}{bn} - C$	$MS_A = \frac{SS_A}{df_A}$
B 因素	$df_b = b-1$	$SS_B = \frac{\sum T_{.j.}^2}{an} - C$	$MS_B = \frac{SS_B}{df_B}$
A×B 互作	$df_{AB} = (a-1)(b-1)$	$SS_{AB} = \frac{\sum T_{ij.}^2}{n} - C - SS_A - SS_B$	$MS_{AB} = \frac{SS_{AB}}{df_{AB}}$
误差	$df_e = ab(n-1)$	$SS_e = SS_T - SS_A - SS_B - SS_{AB}$	$MS_e = \frac{SS_e}{df_e}$
总变异	$df_T = abn-1$	$SS_T = \sum y^2 - C$	

6.4.2 有重复观察值两向分组资料的方差分析

【例6-6】在酒精生产中为了从3种不同原料和3种不同发酵温度中，选出最适宜的条件，设计了一个两因素试验，完全随机设计，每一处理重复4次，并得到下表（表6-17）结果。在这个试验中，温度和原料均为固定因素。试对该资料进行方差分析。

原料种类（A）	重复	温度（B）			$T_{i.}$
		B_1	B_2	B_3	
A_1	1	41	11	6	283
	2	49	13	22	
	3	23	25	26	
	4	25	24	18	
	$T_{ij.}$	138	73	72	
A_2	1	47	43	8	408
	2	59	38	22	
	3	50	33	18	
	4	40	36	14	
	$T_{ij.}$	196	150	62	
A_3	1	43	55	30	473
	2	35	38	33	
	3	53	47	26	
	4	50	44	19	
	$T_{ij.}$	181	184	108	
$T_{.j.}$		515	407	242	T=1164

因素 A	因素 B			总和 $T_{i.}$
	B_1	B_2	B_3	
A_1	138	73	72	283
A_2	196	150	62	408
A_3	181	184	108	473
总和 $T_{.j}$	515	407	242	1164

6.4.2 有重复观察值两向分组资料的方差分析

（1）自由度和平方和的分解

总自由度 $df_T = abn - 1 = 3\times3\times4 - 1 = 35$

处理自由度 $df_t = ab - 1 = 3\times3 - 1 = 8$

A 因素自由度 $df_a = a - 1 = 3 - 1 = 2$

B 因素自由度 $df_b = b - 1 = 3 - 1 = 2$

$A\times B$ 自由度 $df_{AB} = (a-1)(b-1) = (3-1)(3-1) = 4$

误差自由度 $df_e = ab(n-1) = 3\times3(4-1) = 27$

矫正数 $C = \dfrac{T^2}{abn} = \dfrac{1164^2}{36} = 37636.0000$

总平方和 $SS_T = \sum y^2 - C = 41^2 + 11^2 + \cdots + 19^2 - 37636 = 7170.0000$

处理平方和 $SS_t = \dfrac{\sum T_{ij.}^2}{n} - C = \dfrac{138^2 + 73^2 + \cdots + 108^2}{4} - 37636.0000 = 5513.5000$

A 因素平方和 $SS_A = \dfrac{\sum T_{i..}^2}{bn} - C = \dfrac{283^2 + 408^2 + 473^2}{3\times4} - 37636.0000 = 1554.1667$

B 因素平方和 $SS_B = \dfrac{\sum T_{.j.}^2}{an} - C = \dfrac{515^2 + 407^2 + 242^2}{3\times4} - 37636.0000 = 3150.5000$

6.4.2 有重复观察值两向分组资料的方差分析

$A \times B$ 平方和

$$SS_{AB} = \frac{\sum T_{ij.}^2}{n} - C - SS_A - SS_B$$

$$= \frac{138^2 + 73^2 + \cdots + 108^2}{4} - 37636.0000 - 1554.1667 - 3150.5000 = 808.8333$$

误差平方和 $SS_e = SS_T - SS_A - SS_B - SS_{AB} = 7170.0000 - 1554.1667 - 3150.5000 - 808.8333$

（2）计算方差（均方）

$$MS_t = \frac{SS_t}{df_t} = \frac{5513.5000}{8} = 689.1875$$

$$MS_A = \frac{SS_A}{df_A} = \frac{1554.1667}{2} = 777.0833$$

$$MS_B = \frac{SS_B}{df_B} = \frac{3150.5000}{2} = 575.2500$$

$$MS_{AB} = \frac{SS_{AB}}{df_{AB}} = \frac{808.8333}{4} = 202.2083$$

$$MS_e = \frac{SS_e}{df_e} = \frac{1656.5000}{27} = 61.3519$$

6.4.2 有重复观察值两向分组资料的方差分析

（3）列方差分析表，并进行 F 检验

表 6-18 用不同原料与不同温度发酵的酒精产量的方差分析表

变异来源	DF	SS	MS	F	$F_{0.05}$	$F_{0.01}$
处理组合间	8	5513.5000	689.1875	11.23	2.31	3.26
A（原料种类）	2	1554.1667	777.0833	12.67	3.35	5.49
B（温度）	2	3150.5000	575.2500	25.68	3.35	5.49
$A\times B$	4	808.8333	202.2083	3.30	2.73	4.11
误差	27	1656.5000	61.3519			
总变异	35	7170.0000				

A 因素（原料种类）的 $F>F_{0.01(2,27)}=5.49$，表明原料种类对酒精产量差异极显著；B 因素（温度间）的 $F>F_{0.01(2,27)}=5.49$，表明温度对酒精产量差异极显著；原料种类与温度互作（$A\times B$）$F>F_{0.05(4,27)}=2.73$，说明原料种类与温度间存在显著的互作，即说明不同种类的原料获得最高酒精产量的最佳温度是不同的。

6.4.2 有重复观察值两向分组资料的方差分析

（4）多重比较

①不同原料对酒精产量的多重比较，采用 *SSR* 法。因为 *A* 因素（不同原料）各水平的重复数为 *bn*，则 *A* 因素各水平的标准误为：

$$S_{\bar{y}_A} = \sqrt{\frac{MS_e}{bn}} = \sqrt{\frac{61.3519}{3 \times 4}} = 2.2611$$

当 df_e=27，p=2、3，查附表 6，得 α=0.05，α=0.01 的临界 *SSR* 值，用公式 $LSR_\alpha = S_{\bar{y}} \times SSR_\alpha$ 计算出最小显著极差时 *LSR*。*SSR* 值及 *LSR* 值列于表 6-19。

表 6-19 多重比较时的 LSR_α 值计算

秩次距 p	2	3
$SSR_{0.05}$	2.90	3.05
$SSR_{0.01}$	3.91	4.09
$LSR_{0.05}$	6.557	6.896
$LSR_{0.01}$	8.841	9.248

6.4.2 有重复观察值两向分组资料的方差分析

3 种不同原料对酒精产量的多重比较见表 6-20

表 6-20 3 种不同原料对酒精产量的显著性（*SSR* 检验）

原料种类（A）	酒精产量/kg	差异显著性	
		5%	1%
A_3	39.417	a	A
A_2	34.000	a	A
A_1	23.583	b	B

原料种类间多重比较结果解释：从原料种类主效看，原料种类 A_3 酒精产量为最高，极显著高于 A_1；原料种类 A_2 极显著高于 A_1；原料种类 A_3、A_2 间差异不显著。

②不同温度对酒精产量的多重比较，采用 *SSR* 法。因为 *B* 因素（温度）各水平的重复数为 *an*，则 *B* 因素各水平的标准误为：

$$S_{\bar{y}_B}=\sqrt{\frac{MS_e}{an}}=\sqrt{\frac{61.3519}{3\times 4}}=2.2611$$

当 df_e=27，*p*=2、3，查附表 6，得 *α*=0.05，*α*=0.01 的临界 *SSR* 值，用公式 $LSR_\alpha=S_{\bar{y}}\times SSR_\alpha$ 计算出最小显著极差时 *LSR*。*SSR* 值及 *LSR* 值列于表 6-21。

表 6-21 多重比较时的 LSR_α 值计算

秩次距 *p*	2	3
$SSR_{0.05}$	2.90	3.05
$SSR_{0.01}$	3.91	4.09
$LSR_{0.05}$	6.557	6.896
$LSR_{0.01}$	8.841	9.248

温度对酒精产量的多重比较见表 6-22

表 6-22 温度对酒精产量的显著性（*SSR* 检验）

温度（B）	酒精产量/kg	差异显著性	
		5%	1%
B_1	42.917	a	A
B_2	33.917	b	B
B_3	20.167	c	C

温度间多重比较结果解释：从温度主效看，酒精产量以 B_1 最高，B_1、B_2、B_3 间差异极显著。

6.4.2 有重复观察值两向分组资料的方差分析

③各处理组合平均数的比较，采用 *SSR* 法。原料种类与温度的互作显著，说明各处理组合的效应不是各单因素效应的简单相加，而是原料种类效应随温度不同而不同，反之亦然。所以需比较各处理组合的平均数。各处理组合的平均数标准误为：

$$S_{\bar{y}} = \sqrt{\frac{MS_e}{n}} = \sqrt{\frac{61.3519}{4}} = 3.9164$$

当 df_e=27，p=2、3、…、9，查附表 6，得 α=0.05，α=0.01 的临界 *SSR* 值，用公式 $LSR_\alpha = S_{\bar{y}} \times SSR_\alpha$ 计算出最小显著极差时 *LSR*。*SSR* 值及 *LSR* 值列于表 6-23。

表 6-23 多重比较时的 LSR_α 值计算

秩次距 p	2	3	4	5	6	7	8	9
$SSR_{0.05}$	2.90	3.05	3.14	3.21	3.26	3.30	3.33	3.36
$SSR_{0.01}$	3.92	4.09	4.20	4.28	4.35	4.40	4.44	4.48
$LSR_{0.05}$	11.364	11.940	12.311	12.575	12.773	12.928	13.051	13.152
$LSR_{0.01}$	15.346	16.005	16.445	16.768	17.019	17.222	17.390	17.532

6.4.2 有重复观察值两向分组资料的方差分析

表 6-24 处理组合的显著性（*SSR* 检验）

处理	酒精产量/kg	差异显著性	
		5%	1%
A_2B_1	49.000	a	A
A_3B_2	46.000	ab	A
A_3B_1	45.250	ab	A
A_2B_2	37.500	abc	AB
A_1B_1	34.500	bc	AB
A_3B_3	27.000	cd	BC
A_1B_2	18.250	d	C
A_1B_3	18.000	d	C
A_2B_3	15.500	d	C

处理间多重比较结果解释：A_2B_1处理组合的酒精产量最高，极显著高于A_3B_3、A_1B_2、A_1B_3、A_2B_3，显著高于A_1B_1；A_3B_2、A_3B_1处理组合酒精产量极显著高于A_3B_3、A_1B_2、A_1B_3、A_2B_3；A_2B_2、A_1B_1处理组合酒精产量极显著高于A_1B_2、A_1B_3、A_2B_3；A_2B_1、A_3B_2、A_3B_1、A_2B_2间差异不显著；A_3B_2、A_3B_1、A_2B_2、A_1B_1间差异不显著；A_2B_2、A_1B_1、A_3B_3间差异不显著；A_3B_3、A_1B_2、A_1B_3、A_2B_3间差异不显著。

综上所述，最优组合为A_2B_1，即原料种类以A_2和温度B_1组合的酒精产量最高，均值为 49.00。

第六章 方差分析基础

6.5 系统分组资料的方差分析

单向分组资料，若每组又分若干个亚组，每个亚组内又有若干个观测值，则为组内分亚组的单向分组资料，或称**系统分组资料**。

系统分组并不限于组内仅分亚组，亚组内还可分小组，小组内还可分小亚组，……，这种试验称为**巢式试验**（nested experiment）。

系统分组（巢式）设计的基本特征是所分次级之间存在着严格的从属关系，同时由于次级组要适应上一级的需要，因素间不存在交互作用。

最简单的系统分组资料是二级系统分组资料，例如对 l 块土地取样分析，每块地取 m 个样点，每一样点的土样又作 n 次测定。

6.5 系统分组资料的方差分析

1. 二级系统分组资料方差分析的一般思路

设一系统分组资料共有 l 组，每组有 m 个亚组，每个亚组有 n 观测值，共有 lmn 个观测值。

组	亚组	观测值						亚组总和	亚组均数	组总和	组均数
		1	2	…	k	…	n	T_{ij}	$\bar{y}_{ij}$	T_i	$\bar{y}_i$
1	…	…	…	…	…	…	…	…	…	T_1	$\bar{y}_1$
…	…	…	…	…	…	…	…	…	…	…	…
	1	y_{i11}	y_{i12}	…	y_{i1k}	…	y_{i1n}	T_{i1}	$\bar{y}_{i1}$		
	2	y_{i21}	y_{i22}	…	y_{i2k}	…	y_{i2n}	T_{i2}	$\bar{y}_{i2}$		
	…	…	…	…	…	…	…	…	…		
i	j	y_{ij1}	y_{ij2}	…	y_{ijk}	…	y_{ijn}	T_{ij}	$\bar{y}_{ij}$	T_i	$\bar{y}_i$
	…	…	…	…	…	…	…	…	…		
	m	y_{im1}	y_{im2}	…	y_{imk}	…	y_{imn}	T_{im}	$\bar{y}_{im}$		
…	…	…	…	…	…	…	…	…	…	…	…
l										T_l	$\bar{y}_l$
合计								T	$\bar{y}=T/lmn$		

6.5 系统分组资料的方差分析

平方和与自由度的分解

总变异可分解为组间（处理间）变异，组内亚组间变异和亚组内各重复观测值间的变异。

（1）总变异

自由度 $df_T = lmn - 1$

平方和 $SS_T = \sum_{1}^{lmn}(y-\bar{y})^2 = \sum y^2 - \frac{T^2}{lmn} = \sum y^2 - C$

（2）组间（处理间）变异

自由度 $df_t = l - 1$

平方和 $SS_t = mn\sum_{1}^{l}(\bar{y}_i - \bar{y})^2 = \frac{\sum T_i^2}{mn} - C$

（3）组内亚组间变异

自由度 $df_{e_1} = l(m-1)$

平方和 $SS_{e_1} = n\sum_{1}^{l}\sum_{1}^{m}(\bar{y}_{ij} - \bar{y})^2 = \frac{\sum T_{ij}^2}{n} - \frac{\sum T_i^2}{mn}$

（4）试验误差（亚组内各重复观测值间的变异）

自由度 $df_{e_2} = lm(n-1)$

平方和 $SS_{e_2} = \sum_{1}^{l}\sum_{1}^{m}\sum_{1}^{n}(y_{ijk} - \bar{y}_{ij})^2 = \sum y_{ijk}^2 - \frac{\sum T_{ij}^2}{n}$

6.5 系统分组资料的方差分析

扬州大学 YANGZHOU UNIVERSITY

方差分析表

表 6.5.1　二组系统分组资料的方差分析

变异来源	*df*	*SS*	*MS*	*F*
组间（处理间）	*l*-1	SS_t	MS_t	$\frac{MS_t}{MS_{e_1}}$
组内亚组间	*l*(*m*-1)	SS_{e_1}	MS_{e_1}	$\frac{MS_{e_1}}{MS_{e_2}}$
亚组内	*lm*(*n*-1)	SS_{e_2}	MS_{e_2}	
总变异	*lmn*-1	SS_T		

F 测验时 *F* 的计算：当测验各组间有无不同效应时，用 MS_{e_1} 作分母，即 $F = {MS_t}/{MS_{e_1}}$；当测验各亚组间有无不同效应时，用 MS_{e_2} 作分母，即 $F = {MS_{e_1}}/{MS_{e_2}}$。

6.5 系统分组资料的方差分析

2. 实例分析

【例 6-11】 对 A、B、C、D、E 5 个（l=5）水稻品种的干物质积累过程进行了测定，每次测定随机取 2 个(m=2)样点，每个样点取 5 株(n=5)，测定的结果如表 6.5.2，试进行方差分析。

表 6.5.2 5 个杂交水稻品种干物质测定

品种	样点	干物质重量（g/株）					T_{ij}	T_i	$\bar{y}_i$
A	1	7.8	8.9	9.2	11.4	10.5	47.8	99.2	9.92
	2	12.1	10.6	8.7	9.9	10.1	51.4		
B	1	7.4	8.8	8.9	7.8	9.8	42.7	76.4	7.64
	2	6.2	6.6	5.3	7.5	8.1	33.7		
C	1	13.6	14.2	12.1	12.8	13	65.7	133.7	13.37
	2	15.2	15.1	12.3	12.5	12.9	68.0		
D	1	5.8	4.7	6.6	7.4	7.9	32.4	68.8	6.88
	2	6.4	6.8	8.1	7.2	7.9	36.4		
E	1	13.8	15.1	13.4	12.6	16.6	71.5	146.9	14.69
	2	11.7	17.2	15.6	15.1	15.8	75.4		
总和								T=525	

6.5 系统分组资料的方差分析

扬州大学 YANGZHOU UNIVERSITY

2. 实例分析

（1）自由度和平方和的分解

总变异自由度 $df_T = lmn - 1 = 5\times 2\times 5 - 1 = 49$；品种间自由度 $df_t = l - 1 = 5 - 1 = 4$

品种内样点间自由度 $df_{e_1} = l(m-1) = 5(2-1) = 5$；样点内株间自由度 $df_{e_2} = lm(n-1) = 5\times 2(5-1) = 40$

矫正数 $C = \dfrac{T^2}{lmn} = \dfrac{525^2}{50} = 5512.5$

总平方和 $SS_T = \sum y^2 - C = 7.8^2 + 8.9^2 + \cdots + 15.8^2 - 5512.5 = 555.56$

品种间平方和 $SS_t = \dfrac{\sum T_i^2}{mn} - C = \dfrac{99.2^2 + 76.4^2 + \cdots + 146.9^2}{10} - 5512.5 = 474.13$

品种内样点间 $SS_{e_1} = \dfrac{\sum T_{ij}^2}{n} - \dfrac{\sum T_i^2}{mn} = \dfrac{47.8^2 + 51.4^2 + \cdots + 75.4^2}{5} - \dfrac{99.2^2 + 76.4^2 + \cdots + 146.9^2}{10} = 13.05$

样点内株间 $SS_{e_2} = \sum y_{ijk}^2 - \dfrac{\sum T_{ij}^2}{n} = 7.8^2 + 8.9^2 + \cdots + 15.8^2 - \dfrac{47.8^2 + 51.4^2 + \cdots + 75.4^2}{5} = 68.38$

6.5 系统分组资料的方差分析

2. 实例分析

（2）列方差分析表，并进行 F 测验

表 6.5.3　5 个杂交水稻品种干物质测定的方差分析表

变异来源	df	SS	MS	F	$F_{0.05}$	$F_{0.01}$
品种间	4	474.13	118.53	45.43**	5.19	11.3
品种内样点间	5	13.05	2.61	1.53	2.45	3.51
样点内株间	40	68.38	1.71			
总变异	49	555.56				

6.5 系统分组资料的方差分析

扬州大学 YANGZHOU UNIVERSITY

2. 实例分析

(3) 多重比较

$$s_{\bar{y}_i-\bar{y}_j}=\sqrt{\frac{2MS_{e_1}}{mn}}=\sqrt{\frac{2\times 2.61}{2\times 5}}=0.772$$

当自由度 $df=5$ **时，** $t_{0.05}=2.571$ **，** $t_{0.01}=4.032$

$$LSD_{0.05}=t_{0.05}\cdot s_{\bar{y}_i-\bar{y}_j}=2.571\times 0.772=1.985$$

$$LSD_{0.01}=t_{0.01}\cdot s_{\bar{y}_i-\bar{y}_j}=4.032\times 0.772=3.113$$

表 6.5.4　5 个杂交水稻品种干物质积累量的显著性（LSD 测验）

品种	平均值（g/株）	差异显著性	
		0.05	0.01
E	14.69	a	A
C	13.37	a	A
A	9.92	b	B
B	7.64	c	B
D	6.88	c	B

6.5 系统分组资料的方差分析

值得注意的是：在有些教材中指出，若组内亚组间不存在显著差异，在对于组间进行显著性测验时，宜将MS_{e_1}与MS_{e_2}合并，求取合并均方值MS'_e。此时：

$$MS'_e = \frac{SS_{e_1} + SS_{e_2}}{df_{e_1} + df_{e_2}}$$

测验组间的显著性差异，其$F = \frac{MS_t}{MS'_e}$，具有$\nu_1 = l-1$，$\nu_2 = l(mn-1)$。

若用合并均方值MS'_e进行显著性测验，则在多重比较时也需要用合并均方。

第六章 方差分析基础

6.6 变量转换

1、方差分析的基本假定

方差分析也是建立在一些基本假定的基础上。这些基本假定有：

（1）分布的正态性 试验误差是独立的随机变量，并作正态分布的。因为方差分析是假定 k 个样本（处理）从 k 个正态总体中随机抽取的，所以从总体上考虑只有所分析的资料满足正态性条件下才能进行 F 检验。

（2）效应的可加性 处理效应和误差（环境）效应是可加的。这是由于我们据以进行方差分析的模型就是线性可加模型。如果试验资料不具备这一性质，变量的总变异依据变异原因的分解将失去根据，方差分析不能正确进行。

（3）方差的同质性 指各个处理的观测值总体方差σ^2相等。只有这样，才有理由以各个处理均方的合并均方作为检验各处理差异显著性的共同的误差均方。

6.6 变量转换

2、常用变量转换方法

① 平方根转换(square root transformation)

有些取小值的间断性变量，例如单位面积上的虫数或每一视野中的细菌数等，其取值低限为0，高限却可能相当大。这种变量的分布往往不成正态，而其处理平均数往往与其均方成比例。对于这种资料，转换成平方根值往往是很有效的，即：

$$Y' = \sqrt{Y}$$

若原观测值中有为0的数或多数观测值小于10，则把原数据变换成$\sqrt{Y+1}$对于稳定均方，使方差符合一致性的作用更加明显，同时也满足效应可加性和正态性的要求。

6.6 变量转换

② 对数转换(logarithmic transformation)

如果变量的均方是和处理平均数的平方成比例。或者已知处理效应是和处理水平的变化成比例而不是可加的，则宜作对数转换：

$$Y' = \lg Y$$

若Y包含0值，则

$$Y' = \lg(Y+1)$$

一般地说，对数转换对于消弱大变量的作用，要比方根转换更强。例如变量1,10,100，做平方根转换是1, 3.16, 10；做对数转换是0, 1, 2。

6.5 变量转换

③ 反正弦转换(arcsine transformation)

若试验数据是成数或百分数，如发芽率、感病率、死亡率等服从二项分布的百分数资料，方差分析时应作反正弦转换，转换后的数值是以度为单位的角度，因此反正弦转换也称角度转换。转换公式为

$$\theta = \sin^{-1}\sqrt{p}$$

一般，若资料中的百分数介于30%～70%之间时，因资料的分布接近于正态分布，通常数据变换与否对分析结果影响不大；如果资料中的百分数有小于30%或大于70%的，则应对全部百分数进行反正弦转换后，再作方差分析。