第九章　方差分析

在生产过程和科学实验中，我们经常遇到这样的问题：影响产品产量、质量的因素很多.例如，在化工生产中，影响结果的因素有：配方、设备、温度、压力、催化剂、操作人员等.我们需要通过观察或试验来判断哪些因素对产品的产量、质量有显著的影响.方差分析(Analysis of variance)就是用来解决这类问题的一种有效方法.它是在20世纪20年代由英国统计学家费舍尔（Fisher）首先使用到农业试验上去的.后来发现这种方法的应用范围十分广阔，可以成功地应用在试验工作的很多方面.

第一节　单因素试验的方差分析

在试验中，我们将要考察的指标称为试验指标，影响试验指标的条件称为因素.因素可分为两类，一类是人们可以控制的；一类是人们不能控制的.例如，原料成分、反应温度、溶液浓度等是可以控制的，而测量误差、气象条件等一般是难以控制的.以下我们所说的因素都是可控因素，因素所处的状态称为该因素的水平.如果在一项试验中只有一个因素在改变，这样的试验称为单因素试验，如果多于一个因素在改变，就称为多因素试验.

本节通过实例来讨论单因素试验.

1. 数学模型

例 9. 1　某试验室对钢锭模进行选材试验.其方法是将试件加热到700℃后，投入到20℃的水中急冷，这样反复进行到试件断裂为止，试验次数越多，试件质量越好.试验结果如表9-1所示.

试验的目的是确定4种生铁试件的抗热疲劳性能是否有显著差异.

这里，试验的指标是钢锭模的热疲劳值，钢锭模的材质是因素，4种不同的材质表示钢锭模的4个水平，这项试验叫做4水平单因素试验.

表 9-1

试验号	材质分类			
	A1	A2	A3	A4
1	160	158	146	151
2	161	164	155	152
3	165	164	160	153
4	168	170	162	157
5	170	175	164	160
6	172		166	168
7	180		174	
8			182	

例 9. 2　考察一种人造纤维在不同温度的水中浸泡后的缩水率，在40℃，50℃,⋯,90℃的水中分别进行4次试验.得到该种纤维在每次试验中的缩水率（百分比）如表9-2所示.试问浸泡水的温度对缩水率有无显著的影响？

表 9-2 （%）

试验号	温度					
	40℃	50℃	60℃	70℃	80℃	90℃
1	4.3	6.1	10.0	6.5	9.3	9.5
2	7.8	7.3	4.8	8.3	8.7	8.8
3	3.2	4.2	5.4	8.6	7.2	11.4
4	6.5	4.1	9.6	8.2	10.1	7.8

这里试验指标是人造纤维的缩水率，温度是因素，这项试验为 6 水平单因素试验.

单因素试验的一般数学模型为：因素 A 有 s 个水平 A_1，A_2，…，A_s,在水平 $A_j(j=1,2,\cdots,s)$ 下进行 $n_j(n_j\geqslant 2)$ 次独立试验，得到如表 9-3 的结果：

表 9-3

水平 / 观测值	A_1	A_2	…	A_s
	x_{11}	x_{12}	…	x_{1s}
	x_{21}	x_{22}	…	x_{2s}
	… ⋮	…	…	…
	$x_{n_1 1}$	$x_{n_2 2}$	…	$x_{n_s s}$
样本总和	$T_{\cdot 1}$	$T_{\cdot 2}$	…	$T_{\cdot s}$
样本均值	$\overline{x}_{\bullet 1}$	$\overline{x}_{\bullet 2}$	…	$\overline{x}_{\bullet s}$
总体均值	μ_1	μ_2	…	μ_s

假定：各水平 $A_j(j=1,2,\cdots,s)$ 下的样本 $x_{ij}\sim N(\mu_j,\sigma^2),i=1,2,\cdots,n_j,j=1,2,\cdots,s$,且相互独立.

故 $X_{ij}-\mu_j$ 可看成随机误差，它们是试验中无法控制的各种因素所引起的，记 $X_{ij}-\mu_j=\varepsilon_{ij}$,则

$$\begin{cases} x_{ij}=\mu_j+\varepsilon_{ij}, i=1,2,\cdots,n_j;\ j=1,2,\cdots,s, \\ \varepsilon_{ij}\sim N(0,\sigma), \text{各}\varepsilon_{ij}\text{相互独立}, \end{cases} \tag{9-1}$$

其中 μ_j 与 σ^2 均为未知参数.（9-1）式称为单因素试验方差分析的数学模型.

方差分析的任务是对于模型（9-1），检验 s 个总体 $N(\mu_1,\ \sigma^2)$, $N(\mu_2,\ \sigma^2)\cdots,N(\mu_s,\sigma^2)$的均值是否相等，即检验假设：

$$\begin{cases} H_0:\mu_1=\mu_2=\cdots=\mu_s; \\ H_1:\mu_1,\mu_2,\cdots,\mu_s\text{不全相等}. \end{cases} \tag{9-2}$$

为将问题（9-2）写成便于讨论的形式，采用记号

$$\mu=\frac{1}{n}\sum_{j=1}^{s} n_j\mu_j\ ,$$

其中 $n=\sum_{j=1}^{s} n_j$ ，μ 表示 μ_1， $\mu_2,\cdots,\mu_s$ 的加权平均，μ 称为总平均.又记

$$\delta_j=\mu_j-\mu,\qquad j=1,2,\cdots,s,$$

δ_j 表示水平 Aj 下的总体平均值与总平均的差异.习惯上将 δ_j 称为水平 A_j 的效应.利用这些记号，模型（9−1）可改写成：

$$\begin{cases} x_{ij} = \mu + \delta_j + \varepsilon_{ij}, i = 1,2,\cdots,s, \\ \varepsilon_{ij} \sim N\left(0,\sigma^2\right),\ 各\varepsilon_{ij}独立. \end{cases} \quad (9-1)'$$

x_{ij} 可分解成总平均、水平 A_j 的效应及随机误差 3 部分之和,且

$$\sum_{j=1}^{s} n_j \delta_j = 0 .$$

假设（9−2）等价于假设

$$\begin{cases} H_0 : \delta_1 = \delta_2 = \cdots = \delta_s = 0; \\ H_1 : \delta_1, \delta_2, \cdots, \delta_s 不全零. \end{cases} \quad (9.2)'$$

2. 平方和分解

我们寻找适当的统计量，对参数作假设检验.下面从平方和的分解着手，导出假设检验（9-2）′ 的检验统计量.记

$$S_T = \sum_{j=1}^{s}\sum_{i=1}^{n_j} (X_{ij} - \bar{X})^2 , \quad (9-3)$$

这里 $\bar{x} = \frac{1}{n}\sum_{j=1}^{s}\sum_{i=1}^{n_j} X_{ij}$ ，S_T 能反映全部试验数据之间的差异.又称为总变差.

A_j 下的样本均值 $$\bar{X}_{\bullet j} = \frac{1}{n_j}\sum_{i=1}^{n_j} x_{ij} . \quad (9-4)$$

注意到

$$(X_{ij} - \bar{X})^2 = (X_{ij} - \bar{X}_{\bullet j} + \bar{X}_{\bullet j} - \bar{X})^2 = (X_{ij} - \bar{X}_{\bullet j})^2 + (\bar{X}_{\bullet j} - \bar{X})^2 + 2(X_{ij} - \bar{X}_{\bullet j})(\bar{X}_{\bullet j} - \bar{X}) ,$$

而 $$\sum_{j=1}^{s}\sum_{i=1}^{n_j} (X_{ij} - \bar{X}_{\bullet j})(\bar{X}_{\bullet j} - \bar{X}) = \sum_{j=1}^{s} (\bar{X}_{\bullet j} - \bar{X})\left[\sum_{i=1}^{n_j} (X_{ij} - \bar{X}_{\bullet j})\right]$$

$$= \sum_{j=1}^{s} (\bar{X}_{\bullet j} - \bar{X})\left(\sum_{i=1}^{n_j} X_{ij} - n_j \bar{X}_{\bullet j}\right) = 0.$$

记 $$S_E = \sum_{j=1}^{s}\sum_{i=1}^{n_j} (X_{ij} - \bar{X}_{\bullet j})^2 , \quad (9-5)$$

S_E 称为误差平方和;

记 $$S_A = \sum_{j=1}^{s}\sum_{i=1}^{n_j} (\bar{X}_{\bullet j} - \bar{X})^2 = \sum_{j=1}^{s} n_j (\bar{X}_{\bullet j} - \bar{X})^2 , \quad (9-6)$$

S_A 称为因素 A 的效应平方和.于是

$$S_T = S_E + S_A. \quad (9.7)$$

利用 ε_{ij} 可更清楚地看到 S_E, S_A 的含义，记

$$\overline{\varepsilon}=\frac{1}{n}\sum_{j=1}^{s}\sum_{i=1}^{n_j}\varepsilon_{ij}$$

为随机误差的总平均,

$$\overline{\varepsilon}_{\bullet j}=\frac{1}{n_j}\sum_{i=1}^{n_j}\varepsilon_{ij}\,,\qquad j=1,2,\cdots,s.$$

于是

$$S_E=\sum_{j=1}^{s}\sum_{i=1}^{n_j}(X_{ij}-\overline{X}_{\bullet j})^2=\sum_{j=1}^{s}\sum_{i=1}^{n_j}(\varepsilon_{ij}-\overline{\varepsilon}_{\bullet j})^2\,; \qquad (9\text{-}8)$$

$$S_A=\sum_{j=1}^{s}n_j(\overline{X}_{\bullet j}-\overline{X})^2=\sum_{j=1}^{s}n_j(\delta_j+\overline{\varepsilon}_{\bullet j}-\overline{\varepsilon})^2\,. \qquad (9\text{-}9)$$

平方和的分解公式（9−7）说明.总平方和分解成误差平方和与因素 A 的效应平方和.（9-8）式说明 S_E 完全是由随机波动引起的.而（9-9）式说明 S_A 除随机误差外还含有各水平的效应 δ_j，当 δ_j 不全为零时，S_A 主要反映了这些效应的差异.若 H_0 成立，各水平的效应为零，S_A 中也只含随机误差，因而 S_A 与 S_E 相比较相对于某一显著性水平来说不应太大.方差分析的目的是研究 S_A 相对于 S_E 有多大，若 S_A 比 S_E 显著地大，这表明各水平对指标的影响有显著差异.故需研究与 S_A/S_E 有关的统计量.

3. 假设检验问题

当 H_0 成立时，设 $X_{ij}\sim N(\mu,\sigma^2)(i=1,2,\cdots,n_j;\ j=1,2,\cdots,s)$且相互独立，利用抽样分布的有关定理，我们有

$$\frac{S_A}{\sigma^2}\sim\chi^2(s-1), \qquad (9\text{-}10)$$

$$\frac{S_E}{\sigma^2}\sim\chi^2(n-s), \qquad (9\text{-}11)$$

$$F=\frac{(n-s)S_A}{(s-1)S_E}\sim F(s-1,n-s). \qquad (9\text{-}12)$$

于是，对于给定的显著性水平 α $(0<\alpha<1)$,由于

$$P\{F\geqslant F_\alpha(s-1,n-s)\}=\alpha, \qquad (9\text{-}13)$$

由此得检验问题（9. 2）′ 的拒绝域为

$$F\geqslant F_\alpha(s-1,n-s). \qquad (9\text{-}14)$$

由样本值计算 F 的值，若 $F\geqslant F_\alpha$,则拒绝 H_0，即认为水平的改变对指标有显著性的影响；若 $F<F_\alpha$,则接受原假设 H_0，即认为水平的改变对指标无显著影响.

上面的分析结果可排成表 9-4 的形式，称为方差分析表.

表 9-4

方差来源	平方和	自由度	均方和	F 比
因素 A	S_A	s-1	$\overline{S}_A=\frac{S_A}{s-1}$	$F=\overline{S}_A/\overline{S}_E$
误差	S_E	n-s	$\overline{S}_E=\frac{S_E}{n-s}$	
总和	S_T	n-1		

当 $F \geqslant F_{0.05}(s-1,n-s)$时，称为显著，

当 $F \geqslant F_{0.01}(s-1,n-s)$时，称为高度显著.

在实际中，我们可以按以下较简便的公式来计算 S_T，S_A 和 S_E.记

$$T_{\cdot j}=\sum_{i=1}^{n_j} x_{ij}\,, \qquad j=1,2,\cdots,s,$$

$$T_{\cdot\cdot}=\sum_{j=1}^{s}\sum_{i=1}^{n_j} x_{ij}\,,$$

即有

$$\begin{cases} S_T=\sum_{j=1}^{s}\sum_{i=1}^{n_j} x_{ij}^2-n\bar{x}^2=\sum_{j=1}^{s}\sum_{i=1}^{n_j} x_{ij}^2-\dfrac{T_{\bullet\bullet}^2}{n}, \\ S_A=\sum_{j=1}^{s} n_j \bar{x}_{\bullet j}^2-n\bar{x}^2=\sum_{j=1}^{s}\dfrac{T_{\bullet j}^2}{n_j}-\dfrac{T_{\bullet\bullet}^2}{n}, \\ S_E=S_T-S_A. \end{cases} \tag{9-15}$$

例 9.3 如上所述，在例 9. 1 中需检验假设

$$H_0: \mu_1=\mu_2=\mu_3=\mu_4;\quad H_1: \mu_1, \mu_2, \mu_3, \mu_4 \text{不全相等}.$$

给定 $\alpha=0.05$,完成这一假设检验.

解 $s=4,n_1=7,n_2=5,n_3=8,n_4=6,n=26$.

$$S_T=\sum_{j=1}^{s}\sum_{i=1}^{n_j} x_{ij}^2-\frac{T_{\bullet\bullet}^2}{n}=698959-\frac{(4257)^2}{26}=1957.12,$$

$$S_A=\sum_{j=1}^{s}\frac{T_{\bullet j}^2}{n_j}-\frac{T_{\bullet\bullet}^2}{n}=697445.49-\frac{(4257)^2}{26}=443.61,$$

$$S_E=S_T-S_A=1513.51.$$

得方差分析表 9-5.

表 9-5

方差来源	平方和	自由度	均方和	F 比
因素 误差	443.61 1513.51	3 22	147.87 68.80	2.15
总和	1957.12	25		

因

$$F(3,22)=2.15<F_{0.05}(3,22)=3.05.$$

则接受 H_0，即认为 4 种生铁试样的热疲劳性无显著差异.

例 9.4 如上所述，在例 9. 2 中需检验假设

$$H_0: \mu_1=\mu_2=\cdots=\mu_6;\quad H_1: \mu_1, \mu_2,\cdots, \mu_6 \text{不全相等}.$$

试取 $\alpha=0.05$, $\alpha=0.01$，完成这一假设检验.

解 $s=6$, $n_1=n_2=\cdots=n_6=4,n=24$.

$$S_T=\sum_{j=1}^{s}\sum_{i=1}^{n_j} x_{ij}^2-\frac{T_{\bullet\bullet}^2}{n}=112.27,$$

$$S_A=\sum_{j=1}^{s}\frac{T_{\bullet j}^{\ 2}}{n_j}-\frac{T_{\bullet\bullet}^{\ 2}}{n}=56,$$

$$S_E=S_T-S_A=56.27.$$

得方差分析表 9-6.

表 9-6

方差来源	平方和	自由度	均方和	F 比
因素 误差	56 56.27	5 18	11.2 3.126	3.583
总和	112.27	23		

$$F_{0.05}(5,18)=2.77,\quad F_{0.01}(5,18)=4.25.$$

由于 $4.25=F_{0.01}(5,18)>F_A=3.583>F_{0.05}(5,18)=2.77$,

因此浸泡水的温度对缩水率有显著影响，但不能说有高度显著的影响.

本节的方差分析是在这两项假设下，检验各个正态总体均值是否相等.一是正态性假设，假定数据服从正态分布；二是等方差性假设，假定各正态总体方差相等.由大数定律及中心极限定理，以及多年来的方差分析应用，知正态性和等方差性这两项假设是合理的.

第二节　双因素试验的方差分析

进行某一项试验，当影响指标的因素不是一个而是多个时，要分析各因素的作用是否显著，就要用到多因素的方差分析.本节就两个因素的方差分析作一简介.当有两个因素时，除每个因素的影响之外，还有这两个因素的搭配问题.如表 9-7 中的两组试验结果，都有两个因素 A 和 B，每个因素取两个水平.

表 9-7(a)

$B \backslash A$	A_1	A_2
B_1	30	50
B_2	70	90

表 9-7(b)

$B \backslash A$	A_1	A_2
B_1	30	50
B_2	100	80

表 9-7（a）中，无论 B 在什么水平（B_1 还是 B_2），水平 A_2 下的结果总比 A_1 下的高 20；同样地，无论 A 是什么水平，B_2 下的结果总比 B_1 下的高 40.这说明 A 和 B 单独地各自影响结果，互相之间没有作用.

表 9-7(b)中，当 B 为 B_1 时，A_2 下的结果比 A_1 的高，而且当 B 为 B_2 时，A_1 下的结果比 A_2 的高；类似地，当 A 为 A_1 时，B_2 下的结果比 B_1 的高 70，而 A 为 A_2 时，B_2 下的结果比 B_1 的高 30.这表明 A 的作用与 B 所取的水平有关，而 B 的作用也与 A 所取的水平有关.即 A 和 B 不仅各自对结果有影响，而且它们的搭配方式也有影响.我们把这种影响称作因素 A 和 B 的交互作用，记作 $A\times B$.在双因素试验的方差分析中，我们不仅要检验水平 A 和 B 的作用，还要检验它们的交互作用.

1. 双因素等重复试验的方差分析

设有两个因素 A，B 作用于试验的指标，因素 A 有 r 个水平 $A_1,A_2,\cdots,Ar$,因素 B 有 s 个水平 $B_1,B_2,\cdots,B_s$,现对因素 A，B 的水平的每对组合(A_i,B_j),$i=1,2,\cdots,r$；$j=1,2,\cdots,s$ 都作 $t(t\geqslant 2)$次试

验（称为等重复试验），得到如表 9-8 的结果：

表 9-8

因素 B \ 因素 A	B_1	B_2	…	B_s
A_1	$X_{111},X_{112}\cdots,X_{11t}$	$X_{121},X_{122}\cdots,X_{12t}$	…	$X_{1s1},X_{1s2}\cdots,X_{1st}$
A_2	$X_{211},X_{212}\cdots,X_{21t}$	$X_{221},X_{222}\cdots,X_{22t}$	…	$X_{2s1},X_{2s2},\cdots,X_{2st}$
⋯	⋯	⋯	⋯	⋯
A_r	$X_{r11},X_{r12}\cdots,X_{r1t}$	$X_{r21},X_{r22}\cdots,X_{r2t}$	…	$X_{rs1},X_{rs2}\cdots,X_{rst}$

设 $X_{ijk}\sim N(\mu_{ij},\sigma^2)$, $i=1,2,\cdots,r$； $j=1,2,\cdots,s$； $k=1,2,\cdots,t$,各 x_{ijk} 独立.这里 μ_{ij}, σ^2 均为未知参数.或写为

$$\begin{cases} X_{ijk}=\mu_{ij}+\varepsilon_{ijk}, j=1,2,\cdots,r; j=1,2,\cdots,s, \\ \varepsilon_{ijk}\sim N(0,\sigma^2), k=1,2,\cdots,t, \\ \text{各}\,\varepsilon_{ijk}\,\text{相互独立}. \end{cases} \tag{9-16}$$

记

$$\mu=\frac{1}{rs}\sum_{i=1}^{r}\sum_{j=1}^{s}\mu_{ij},,\quad \mu_{i\bullet}=\frac{1}{s}\sum_{j=1}^{s}\mu_{ij},\ i=1,2,\cdots,r,$$

$$\mu_{\bullet j}=\frac{1}{r}\sum_{i=1}^{r}\mu_{ij},\quad j=1,2,\cdots,s,$$

$$\alpha_i=\mu_{i\bullet}-\mu,,\quad i=1,2,\ldots,r,\quad \beta_j=\mu_{\bullet j}-\mu,\quad j=1,2,\ldots,s,$$

$$\gamma_{ij}=\mu_{ij}-\mu_{i\bullet}-\mu_{\bullet j}+\mu.$$

于是 $$\mu_{ij}=\mu+\alpha_i+\beta_j+\gamma_{ij}. \tag{9-17}$$

称 μ 为总平均， α_i 为水平 A_i 的效应， β_j 为水平 B_j 的效应， γ_{ij} 为水平 A_i 和水平 B_j 的交互效应，这是由 A_i,B_j 搭配起来联合作用而引起的.

易知

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

$$\sum_{i=1}^{r}\gamma_{ij}=0,\quad j=1,2,\cdots,s,$$

$$\sum_{j=1}^{s}\gamma_{ij}=0,\quad i=1,2,\cdots,r,$$

这样（9-16）式可写成

$$
\begin{cases}
X_{ijk} = \mu + \alpha_i + \beta_j + \gamma_{ij} + \varepsilon_{ijk}, \\
\sum_{i=1}^{r} \alpha_i = 0, \sum_{j=1}^{s} \beta_j = 0, \sum_{i=1}^{r} \gamma_{ij} = 0, \sum_{j=1}^{s} \gamma_{ij} = 0, \\
\varepsilon_{ijk} \sim N(0, \sigma^2), i = 1, 2, \cdots, r; j = 1, 2, \cdots, s; k = 1, 2, \cdots, t, \\
\text{各}\varepsilon_{ijk}\text{相互独立}.
\end{cases} \tag{9-18}
$$

其中 μ, α_i, β_j, γ_{ij} 及 σ^2 都为未知参数.

（9-18）式就是我们所要研究的双因素试验方差分析的数学模型.我们要检验因素 A，B 及交互作用 $A\times B$ 是否显著.要检验以下 3 个假设：

$$
\begin{cases}
H_{01} : \alpha_1 = \alpha_2 = \cdots = \alpha_r = 0, \\
H_{11} : \alpha_1, \alpha_2, \cdots = \alpha_r \text{不全为零}.
\end{cases}
$$

$$
\begin{cases}
H_{02} : \beta_1 = \beta_2 = \cdots = \beta_s = 0, \\
H_{12} : \beta_1, \beta_2, \cdots = \beta_s \text{不全为零}.
\end{cases}
$$

$$
\begin{cases}
H_{03} : \gamma_{11} = \gamma_{12} = \cdots = \gamma_{rs} = 0, \\
H_{13} : \gamma_{11}, \gamma_{12}, \cdots = \gamma_{rs} \text{不全为零}.
\end{cases}
$$

类似于单因素情况，对这些问题的检验方法也是建立在平方和分解上的.记

$$
\overline{X} = \frac{1}{rst} \sum_{i=1}^{r} \sum_{j=1}^{s} \sum_{k=1}^{t} x_{ijk},
$$

$$
\overline{X}_{ij\bullet} = \frac{1}{t} \sum_{k=1}^{t} x_{ijk}, \quad i=1,2,\cdots,r; \quad j=1,2,\cdots,s,
$$

$$
\overline{X}_{i\bullet\bullet} = \frac{1}{st} \sum_{j=1}^{s} \sum_{k=1}^{t} X_{ijk}, \quad i=1,2,\cdots,r,
$$

$$
\overline{X}_{\bullet j\bullet} = \frac{1}{rt} \sum_{i=1}^{r} \sum_{k=1}^{t} X_{ijk}, \quad j=1,2,\cdots,s,
$$

$$
S_T = \sum_{i=1}^{r} \sum_{j=1}^{s} \sum_{k=1}^{t} (\mathrm{X}_{ijk} - \overline{X})^2 .
$$

不难验证 $\overline{\mathrm{X}}, \overline{X}_{i\bullet\bullet}, \overline{X}_{\bullet j\bullet}, \overline{X}_{ij\bullet}$ 分别是 μ, $\mu_{i\cdot}$, $\mu_{\cdot j}$, μ_{ij} 的无偏估计.

由 $X_{ijk} - \overline{X} = (X_{ijk} - \overline{X}_{ij\bullet}) + (\overline{X}_{i\bullet\bullet} - \overline{X}) + (\overline{X}_{\bullet j\bullet} - \overline{X}) + (\overline{X}_{ij\bullet} - \overline{X}_{i\bullet\bullet} - \overline{X}_{\bullet j\bullet} + \overline{X})$，

$1\leqslant i\leqslant r, 1\leqslant j\leqslant s, 1\leqslant k\leqslant t$

得平方和的分解式：

$$
S_T = S_E + S_A + S_B + S_{A\times B}, \tag{9-19}
$$

其中

$$S_E=\sum_{i=1}^{r}\sum_{j=1}^{s}\sum_{k=1}^{t}(X_{ijk}-\bar{X}_{ij\bullet})^2,$$

$$S_A=st\sum_{i=1}^{r}(\bar{X}_{i\bullet\bullet}-\bar{X})2,$$

$$S_B=rt\sum_{j=1}^{s}(\bar{X}_{\bullet j\bullet}-\bar{X})^2,$$

$$S_{A\times B}=t\sum_{i=1}^{r}\sum_{j=1}^{s}(\bar{X}_{ij\bullet}-\bar{X}_{i\bullet\bullet}-\bar{X}_{\bullet j\bullet}+\bar{X})^2.$$

S_E称为误差平方和，S_A，S_B分别称为因素 A，B 的效应平方和，$S_{A\times B}$称为 A，B 交互效应平方和.

当 H_{01}: $\alpha_1=\alpha_2=\cdots=\alpha_r=0$ 为真时，

$$F_A=\frac{S_A}{(r-1)}\Big/\frac{S_E}{[rs(t-1)]}\sim F(r-1,rs(t-1));$$

当假设 H_{02} 为真时，

$$F_B=\frac{S_B}{(s-1)}\Big/\frac{S_E}{[rs(t-1)]}\sim F(s-1,rs(t-1));$$

当假设 H_{03} 为真时，

$$F_{A\times B}=\frac{S_{A\times B}}{(r-1)(s-1)}\Big/\frac{S_E}{[rs(t-1)]}\sim F((r-1)(s-1),rs(t-1)).$$

当给定显著性水平 α 后，假设 H_{01}，H_{02}，H_{03} 的拒绝域分别为：

$$\begin{cases}F_A\geq F_\alpha(r-1,rs(t-1));\\F_B\geq F_\alpha(s-1,rs(t-1));\\F_{A\times B}\geq F_\alpha(r-1)(s-1),rs(t-1)).\end{cases}\tag{9-20}$$

经过上面的分析和计算，可得出双因素试验的方差分析表 9-9.

表 9-9

方差来源	平方和	自由度	均方和	F 比
因素 A	S_A	r-1	$\bar{S}_A=\frac{S_A}{r-1}$	$F_A=\frac{\bar{S}_A}{\bar{S}_E}$
因素 B	S_B	s-1	$\bar{S}_B=\frac{S_B}{s-1}$	$F_B=\frac{\bar{S}_B}{\bar{S}_E}$
交互作用	$S_{A\times B}$	(r-1)(s-1)	$\bar{S}_{A\times B}=\frac{S_{A\times B}}{(r-1)(s-1)}$	$F_{A\times B}=\frac{\bar{S}_{A\times B}}{\bar{S}_E}$

误差	S_E	$rs(t-1)$	$\overline{S}_E = \dfrac{S_E}{rs(t-1)}$	
总和	S_T	$rst-1$		

在实际中，与单因素方差分析类似可按以下较简便的公式来计算 S_T，S_A，S_B，$S_{A\times B}$,S_E.

记 $T_{\cdots}=\sum_{i=1}^{r}\sum_{j=1}^{s}\sum_{k=1}^{t}X_{ijk}$,

$T_{ij\cdot}=\sum_{k=1}^{t}X_{ijk}$, $i=1,2,\cdots,r$; $j=1,2,\cdots,s$,

$T_{i\cdot\cdot}=\sum_{j=1}^{s}\sum_{k=1}^{t}X_{ijk}$, $i=1,2,\cdots,r$,

$T_{\cdot j\cdot}=\sum_{i=1}^{r}\sum_{k=1}^{t}X_{ijk}$, $j=1,2,\cdots,s$,

即有

$$\begin{cases} S_T=\sum_{i=1}^{r}\sum_{j=1}^{s}\sum_{k=1}^{t}X_{ijk}^{\ 2}-\dfrac{T_{\bullet\bullet\bullet}^{\ 2}}{rst}, \\ S_A=\dfrac{1}{st}\sum_{i=1}^{r}T_{i\bullet\bullet}^{\ 2}-\dfrac{T_{\bullet\bullet\bullet}^{\ 2}}{rst}, \\ S_B=\dfrac{1}{rt}\sum_{j=1}^{s}T_{\bullet j\bullet}^{\ 2}-\dfrac{T_{\bullet\bullet\bullet}^{\ 2}}{rst}, \\ S_{A\times B}=\dfrac{1}{t}\sum_{i=1}^{r}\sum_{j=1}^{s}T_{ij\bullet}^{\ 2}-\dfrac{T_{\bullet\bullet\bullet}^{\ 2}}{rst}-S_A-S_B, \\ S_E=S_T-S_A-S_B-S_{A\times B}. \end{cases} \qquad (9\text{-}21)$$

例 9. 5 用不同的生产方法（不同的硫化时间和不同的加速剂）制造的硬橡胶的抗牵拉强度（以 kg・cm^{-2} 为单位）的观察数据如表 9-10 所示. 试在显著水平 0. 10 下分析不同的硫化时间（A），加速剂（B）以及它们的交互作用（A×B）对抗牵拉强度有无显著影响.

表 9-10

140℃下硫化时间（秒）	加速剂		
	甲	乙	丙
40	39，36	43，37	37，41
60	41，35	42，39	39，40
80	40，30	43，36	36，38

解 按题意，需检验假设 H_{01},H_{02},H_{03}.

r=s=3, t=2，

$T_{\cdots}$,$T_{ij\cdot}$,$T_{i\cdot\cdot}$,$T_{\cdot j\cdot}$ 的计算如表 9-11.

表 9-11

加速剂 $T_{ij\cdot}$ 硫化时间	甲	乙	丙	$T_{i\cdot\cdot}$
40	75	80	78	233
60	76	81	79	236
80	70	79	74	223
$T_{\cdot j\cdot}$	221	240	231	692

$$S_T=\sum_{i=1}^{r}\sum_{j=1}^{s}\sum_{k=1}^{t}x_{ijk}{}^2-\frac{T_{\bullet\bullet\bullet}{}^2}{rst},=178.44,$$

$$S_A=\frac{1}{st}\sum_{i=1}^{r}T_{i\bullet\bullet}{}^2-\frac{T_{\bullet\bullet\bullet}{}^2}{rst}=15.44,$$

$$S_B=\frac{1}{rt}\sum_{j=1}^{s}T_{\bullet j\bullet}{}^2-\frac{T_{\bullet\bullet\bullet}{}^2}{rst}=30.\ 11,$$

$$S_{A\times B}=\frac{1}{t}\sum_{i=1}^{r}\sum_{j=1}^{s}T_{ij\bullet}{}^2-\frac{T_{\bullet\bullet\bullet}{}^2}{rst}-S_A-S_B\quad=2.89,$$

$$S_E=S_T\text{-}S_A\text{-}S_B\text{-}S_{A\times B}=130,$$

得方差分析表 9-12.

表 9-12

方差来源	平方和	自由度	均方和	F 比
因素 A（硫化时间）	15.44	2	7.72	F_A=0.53
因素 B（加速剂）	30.11	2	15.56	F_B=1.04
交互作用 $A\times B$	2.89	4	0.7225	$F_{A\times B}$=0.05
误差	130	9	14.44	
总和	178.44			

由于 $F_{0.10}(2,9)=3.01>F_A$,$F_{0.10}(2,9)>F_B$,$F_{0.10}(4,9)=2.69>F_{A\times B}$,因而接受假设 H_{01},H_{02},H_{03},即硫化时间、加速剂以及它们的交互作用对硬橡胶的抗牵拉强度的影响不显著.

2. 双因素无重复试验的方差分析

在双因素试验中，如果对每一对水平的组合（A_i,B_j）只做一次试验，即不重复试验，所得结果如表 9-13 所示.

表 9-13

因素 B 因素 A	B_1	B_2	…	B_s
A_1	x_{11}	x_{12}	…	x_{1s}
A_2	x_{21}	x_{22}	…	x_{2s}
⋯	⋯	⋯	⋯	⋯
A_r	x_{r1}	x_{r2}	…	x_{rs}

这时 $\overline{X}_{ij\bullet}=X_{ijk}$,$S_E=0$,$S_E$ 的自由度为 0，故不能利用双因素等重复试验中的公式进行方差分析.

但是，如果我们认为 A，B 两因素无交互作用，或已知交互作用对试验指标影响很小，则可

将 $S_{A\times B}$ 取作 S_E，仍可利用等重复的双因素试验对因素 A，B 进行方差分析.对这种情况下的数学模型及统计分析表示如下：

由（9-18）式,

$$\begin{cases} X_{ij}=\mu+\alpha_i+\beta_j+\varepsilon_{ij}, \\ \sum_{i=1}^{r}\alpha_i=0, \sum_{j=1}^{s}\beta_j=0, \\ \varepsilon_{ij}\sim N(0,\sigma^2), i=1,2,\cdots,r; j=1,2,\cdots,s, \\ \text{各}\varepsilon_{ijk}\text{独立}. \end{cases} \tag{9-22}$$

要检验的假设有以下两个：

$$\begin{cases} H_{01}:\alpha_1=\alpha_2=\cdots=\alpha_r=0, \\ H_{11}:\alpha_1,\alpha_2,\cdots=\alpha_r\text{不全为零}. \end{cases}$$

$$\begin{cases} H_{02}:\beta_1=\beta_2=\cdots=\beta_s=0, \\ H_{12}:\beta_1,\beta_2,\cdots=\beta_s\text{不全为零}. \end{cases}$$

记 $$\bar{X}=\frac{1}{rs}\sum_{i=1}^{r}\sum_{j=1}^{s}X_{ij}, \bar{X}_{i\bullet}=\frac{1}{s}\sum_{j=1}^{s}X_{ij}, \bar{X}_{\bullet j}=\frac{1}{r}\sum_{i=1}^{r}X_{ij},$$

平方和分解公式为：

$$S_T=S_A+S_B+S_E, \tag{9-23}$$

其中 $$S_T=\sum_{i=1}^{r}\sum_{j=1}^{s}(X_{ij}-\bar{X})^2,\quad S_A=s\sum_{j=1}^{s}(\bar{X}_{i\bullet}-\bar{X})^2,$$

$$S_B=r\sum_{j=1}^{s}(\bar{X}_{\bullet j}-\bar{X})^2,\quad S_E=\sum_{i=1}^{r}\sum_{j=1}^{s}(X_{ij}-\bar{X}_{i\bullet}-\bar{X}_{\bullet j}+\bar{X})^2,$$

分别为总平方和、因素 A，B 的效应平方和和误差平方和.

取显著性水平为 α,当 H_{01} 成立时，

$$F_A=\frac{(s-1)S_A}{S_E}\sim F((r-1),(r-1)(s-1)),$$

H_{01} 拒绝域为

$$F_A\geqslant F_\alpha((r-1),(r-1)(s-1)). \tag{9-24}$$

当 H_{02} 成立时，

$$F_B=\frac{(r-1)S_B}{S_E}\sim F((s-1),(r-1)(s-1)),$$

H_{02} 拒绝域为

$$F_B\geqslant F_\alpha((s-1),(r-1)(s-1)). \tag{9-25}$$

得方差分析表 9-14.

表 9-14

方差来源	平方和	自由度	均方和	F 比
因素 A 因素 B 误差	S_A SB S_E	r-1 s-1 (r-1)(s-1)	$\bar{S}_A=\dfrac{S_A}{r-1}$ $\bar{S}_B=\dfrac{S_B}{s-1}$ $\bar{S}_E=\dfrac{S_E}{(r-1)(s-1)}$	$F_A=\bar{S}_A/\bar{S}_E$ $F_B=\bar{S}_B/\bar{S}_E$
总和	S_T	rs-1		

例 9. 6　测试某种钢不同含铜量在各种温度下的冲击值（单位：kg・m・cm^{-1}），表 9-15 列出了试验的数据(冲击值)，问试验温度、含铜量对钢的冲击值的影响是否显著？（α=0.01）

表 9-15

铜含量 / 冲击值 / 试验温度	0.2%	0.4%	0.8%
20℃	10.6	11.6	14.5
0℃	7.0	11.1	13.3
-20℃	4.2	6.8	11.5
-40℃	4.2	6.3	8.7

解　由已知，r=4,s=3,需检验假设 H_{01},H_{02}，经计算得方差分析表 9-16.

表 9-16

方差来源	平方和	自由度	均方和	F 比
温度作用	64.58	3	21.53	23.79
铜含量作用	60.74	2	30.37	33.56
试验误差	5.43	6	0.905	
总和	130.75	11		

由于 $F_{0.01}(3,6)=9.78<F_A$,拒绝 H_{01}.

$F_{0.01}$（2,6）=10.92<F_B,拒绝 H_{02}.

检验结果表明，试验温度、含铜量对钢冲击值的影响是显著的.

第三节　正交试验设计及其方差分析

前两节讨论的单因素试验和双因素试验均属于全面试验(即每一个因素的各种水平的相互搭配都要进行试验）.在工农业生产和科学实验中，为改革旧工艺，寻求最优生产条件等，经常要做许多试验，而影响这些试验结果的因素很多，我们把含有两个以上因素的试验称为多因素试验.前两节讨论的单因素试验和双因素试验均属于全面试验（即每一个因素的各种水平的相互搭配都要进行试验），多因素试验由于要考虑的因素较多，当每个因素的水平数

较大时，若进行全面试验，则试验次数将会更大.因此，对于多因素试验，存在一个如何安排好试验的问题.正交试验设计是研究和处理多因素试验的一种科学方法，它利用一套现存规格化的表——正交表，来安排试验，通过少量的试验，获得满意的试验结果.

1. 正交试验设计的基本方法

正交试验设计包含两个内容：（1）怎样安排试验方案；（2）如何分析试验结果.先介绍正交表.

先介绍正交表.

正交表是预先编制好的一种表格.比如表 9-17 即为正交表 $L_4(2^3)$,其中字母 L 表示正交，它的 3 个数字有 3 种不同的含义：

表 9-17

列号 / 试验号	1	2	3
1	1	1	1
2	1	2	2
3	2	1	2
4	2	2	1

(1) L_4（2^3）表的结构：有 4 行、3 列，表中出现 2 个反映水平的数码 1，2.

列数

↓

L_4 （2^3）

↑ ↑

行数 水平数

（2） L_4（2^3）表的用法：做 4 次试验，最多可安排 2 水平的因素 3 个.

最多能安排的因素数

↓

L_4 (2^3)

↑ ↑

试验次数 水平数

(3) L_4（2^3）表的效率：3 个 2 水平的因素.它的全面试验数为 $2^3=8$ 次，使用正交表只需从 8 次试验中选出 4 次来做试验，效率是高的.

L_4 (2^3)

↑ ↑

实际试验数 理论上的试验数

正交表的特点：

（1） 表中任一列，不同数字出现的次数相同.如正交表 $L_4(2^3)$中，数字 1，2 在每列中均出现 2 次.

（2） 表中任两列，其横向形成的有序数对出现的次数相同.如表 L_4（2^3）中任意两列，数字 1，2 间的搭配是均衡的.

凡满足上述两性质的表都称为正交表(Orthogonal table).

常用的正交表有 L_9（3^4），L_8（2^7），L_{16}（4^5）等，见附表 7.用正交表来安排试验的方法，就叫正交试验设计.一般正交表 L_p（n^m）中，$p=m(n-1)+1$.下面通过实例来说明如何用正交表来安排试验.

例 9. 7 提高某化工产品转化率的试验.某种化工产品的转化率可能与反应温度 A，反应时间 B，某两种原料之配比 C 和真空度 D 有关.为了寻找最优的生产条件，因此考虑对 A，B，C，D 这 4 个因素进行试验.根据以往的经验，确定各个因素的 3 个不同水平，如表 9-18 所示.

表 9-18

水平 因素	1	2	3
A:反应温度（℃）	60	70	80
B:反应时间（小时）	2.5	3.0	3.5
C:原料配比	1.1∶1	1.15∶1	1.2∶1
D:真空度（毫米汞柱）	500	550	600

分析各因素对产品的转化率是否产生显著影响，并指出最好生产条件.

解 本题是 4 因素 3 水平，选用正交表 $L_9(3^4)$.

表 9-19

列号 水平 试验号	A 1	B 2	C 3	D 4
1	1	1	1	1
2	1	2	2	2
3	1	3	3	3
4	2	1	2	3
5	2	2	3	1
6	2	3	1	2
7	3	1	3	2
8	3	2	1	3
9	3	3	2	1

把表头上各因素相应的水平任意给一个水平号. 本例的水平编号就采用表 9-18 的形式；将各因素的诸水平所表示的实际状态或条件代入正交表中，得到 9 个试验方案，如表 9-20 所示.

表 9-20

列号 水平 试验号	A 1	B 2	C 3	D 4
1	1(60)	1(2.5)	1(1.1:1)	1(500)
2	1	2(3.0)	2(1.15:1)	2(550)
3	1	3(3.5)	3(1.2:1)	3(600)
4	2(70)	1	2	3
5	2	2	3	1
6	2	3	1	2
7	3(80)	1	3	2
8	3	2	1	3
9	3	3	2	1

从表 9-20 看出，第一行是 1 号试验，其试验条件是：

反应温度为 60℃，反应时间为 2. 5 小时，原料配比为 1.1∶1，真空度为 500mmHg，记

作 $A_1B_1C_1D_1$.依此类推，第 9 号试验条件是 $A_3B_3C_2D_1$.

由此可见，因素和水平可以任意排，但一经排定，试验条件也就完全确定.按正交试验表 9-20 安排试验，试验的结果依次记于试验方案右侧，见表 9-21.

表 9-21

列号 / 水平 / 试验号	*A*	*B*	*C*	*D*	试验结果（%）
1	1(60)	1(2.5)	1(1.1:1)	1(500)	38
2	1	2(3.0)	2(1.15:1)	2(550)	37
3	1	3(3.5)	3(1.2:1)	3(600)	76
4	2(70)	1	2	3	51
5	2	2	3	1	50
6	2	3	1	2	82
7	3(80)	1	3	2	44
8	3	2	1	3	55
9	3	3	2	1	86

2. 试验结果的直观分析

正交试验设计的直观分析就是要通过计算，将各因素、水平对试验结果指标的影响大小，通过极差分析，综合比较，以确定最优化试验方案的方法.有时也称为极差分析法.

例 9. 7 中试验结果转化率列在表 9-21 中，在 9 次试验中，以第 9 次试验的指标 86 为最高，其生产条件是 $A_3B_3C_2D_1$.由于全面搭配试验有 81 种，现只做了 9 次.9 次试验中最好的结果是否一定是全面搭配试验中最好的结果呢？还需进一步分析.

（1） 极差计算

在代表因素 A 的表 9-21 的第 1 列中，将与水平“1”相对应的第 1，2，3 号 3 个试验结果相加，记作 T_{11}，求得 T_{11}=151.同样，将第 1 列中与水平“2”对应的第 4，5，6 号试验结果相加，记作 T_{21}，求得 T_{21}=183.

一般地，定义 T_{ij} 为表 9-21 的第 j 列中，与水平 i 对应的各次试验结果之和(i=1,2,3; j=1,2,3,4).记 T 为 9 次试验结果的总和，R_j 为第 j 列的 3 个 T_{ij} 中最大值与最小值之差，称为极差.

显然 $T=\sum_{i=1}^{3} T_{ij}$，j=1,2,3,4.

此处 T_{11} 大致反映了 A_1 对试验结果的影响，

T_{21} 大致反映了 A_2 对试验结果的影响，

T_{31} 大致反映了 A_3 对试验结果的影响，

T_{12}，T_{22} 和 T_{32} 分别反映了 B_1，B_2，B_3 对试验结果的影响，

T_{13}，T_{23} 和 T_{33} 分别反映了 C_1，C_2，C_3 对试验结果的影响，

T_{14}，T_{24} 和 T_{34} 分别反映了 D_1，D_2，D_3 对试验结果的影响.

R_j 反映了第 j 列因素的水平改变对试验结果的影响大小，R_j 越大反映第 j 列因素影响越大.上述结果列于表 9-22 中.

表 9-22

T_{1j}	151	133	175	174	T=519
T_{2j}	183	142	174	163	
T_{3j}	185	244	170	182	

R_j	34 111 5 19	

(2) 极差分析(Analysis of range)

由极差大小顺序排出因素的主次顺序：

主→次

B；A，D；C

这里，R_j 值相近的两因素间用“，”号隔开，而 R_j 值相差较大的两因素间用“；”号隔开.由此看出，特别要求在生产过程中控制好因素 B，即反应时间.其次是要考虑因素 A 和 D，即要控制好反应温度和真空度.至于原料配比就不那么重要了.

选择较好的因素水平搭配与所要求的指标有关.若要求指标越大越好，则应选取指标大的水平.反之，若希望指标越小越好，应选取指标小的水平.例 9. 7 中，希望转化率越高越好，所以应在第 1 列选最大的 T_{31}=185；即取水平 A_3，同理可选 $B_3C_1D_3$.故例 9. 7 中较好的因素水平搭配是 $A_3B_3C_1D_3$.

例 9. 8 某试验被考察的因素有 5 个：A，B，C，D，E.每个因素有两个水平.选用正交表 $L_8(2^7)$，现分别把 A，B，C，D，E 安排在表 $L_8(2^7)$的第 1，2，4，5，7 列上，空出第 3，6 列，仿例 9. 7 做法，按方案试验.记下试验结果，进行极差计算，得表 9-23.

表 9-23

列号 / 水平 / 试验号	A 1	B 2	 3	C 4	D 5	 6	E 7	试验结果
1	1	1	1	1	1	1	1	14
2	1	1	1	2	2	2	2	13
3	1	2	2	1	1	2	2	17
4	1	2	2	2	2	1	1	17
5	2	1	2	1	2	1	2	8
6	2	1	2	2	1	2	1	10
7	2	2	1	1	2	2	1	11
8	2	2	1	2	1	1	2	15
T_{1j}	61	45	53	50	56	54	52	T=105
T_{2j}	44	60	52	55	49	51	53	
R_j	17	15	1	5	7	3	1	

试验目的要找出试验结果最小的工艺条件及因素影响的主次顺序.从表 9-23 的极差 R_j 的大小顺序排出因素的主次顺序为

主 → 次

A，B；D；C，E

最优工艺条件为 $A_2B_1C_1D_2E_1$.

表 9-23 中因没有安排因素而空出了第 3，6 列.从理论上说，这两列的极差 R_j 应为 0，但因存有随机误差，这两个空列的极差值实际上是相当小的.

3. 方差分析

正交试验设计的极差分析简便易行，计算量小，也较直观，但极差分析精度较差，判断因素的作用时缺乏一个定量的标准.这些问题要用方差分析解决.

设有一试验，使用正交表 $L_p(n^m)$,试验的 p 个结果为 $y_1,y_2,\cdots,y_p$,记

$$T=\sum_{i=1}^{p} y_i\,, \qquad \bar{y}=\frac{1}{p}\sum_{i=1}^{p} y_i=\frac{T}{p}\,,$$

$$S_T=\sum_{i=1}^{p}(y_i-\overline{y})^2$$

为试验的 p 个结果的总变差；

$$S_j=r\sum_{i=1}^{n}\left(\frac{T_{ij}}{r}-\frac{T}{p}\right)^2=\frac{1}{r}\sum_{i=1}^{n}{T_{ij}}^2-\frac{T^2}{p}$$

为第 j 列上安排因素的变差平方和，其中 $r=p/n$.可证明

$$S_T=\sum_{j=1}^{m}S_i$$

即总变差为各列变差平方和之和，且 S_T 的自由度为 p-1，S_j 的自由度为 n-1.当正交表的所有列没被排满因素时，即有空列时，所有空列的 S_j 之和就是误差的变差平方和 S_e，这时 S_e 的自由度 f_e 也为这些空列自由度之和.当正交表的所有列都排有因素时，即无空列时，取 S_j 中的最小值作为误差的变差平方和 S_e.

从以上分析知，在使用正交表 $L_p(n^m)$的正交试验方差分析中，对正交表所安排的因素选用的统计量为：

$$F=\frac{S_j}{n-1}\Big/\frac{S_e}{f_e}.$$

当因素作用不显著时，

$$F\sim F(n-1,f_e),$$

其中第 j 列安排的是被检因素.

在实际应用时，先求出各列的 $S_j/(n-1)$及 S_e/f_e,若某个 $S_j/(n-1)$比 S_e/f_e 还小时，则这第 j 列就可当作误差列并入 S_e 中去，这样使误差 S_e 的自由度增大，在作 F 检验时会更灵敏，将所有可当作误差列的 S_j 全并入 S_e 后得新的误差变差平方和，记为 $S_e^{\triangle}$,其相应的自由度为 $f_e^{\triangle}$,这时选用统计量

$$F=\frac{S_j}{n-1}\Big/\frac{S_e^{\triangle}}{f_e^{\triangle}}\sim F(n-1,f_e^{\triangle}).$$

例 9. 9 对例 9. 8 的表 9-23 作方差分析.

解 由表 9-23 的最后一行的极差值 R_j，利用公式 $S_j=\frac{1}{r}\sum_{i=1}^{n}{T_{ij}}^2-\frac{T^2}{p}$,得表 9-24.

表 9-24

	A 1	*B* 2	3	*C* 4	*D* 5	6	*E* 7	
R_j	17	15	1	5	7	3	1	
S_j	36.125	28.125	0.125	3.125	6.125	1.125	0.125	S_T=74.875

表 9-24 中,第 3，6 列为空列，因此 $S_e=S_3+S_6=1.250$,其中 f_e=1+1=2,所以 S_e/f_e=0.625,而第 7 列的 S_7=0.125，S_7/f_7=0.1251=0.125 比 S_e/f_e 小，故将它并入误差.

$S_e^{\triangle}=S_e+S_7=1.375$,$f_e^{\triangle}=3$.整理成方差分析表如表 9-25 所示.

表 9-25

方差来源	S_j	f_j	$\dfrac{S_j}{f_j}$	$F=\dfrac{S_j}{fj}\Big/\dfrac{S_e^{\Delta}}{f_e^{\Delta}}$	显著性
A	36.125	1	36.125	78.818	
B	28.125	1	28.125	61.364	
C	3.125	1	3.125	6.818	
D	6.125	1	6.125	13.364	
E^{Δ}	0.125	1	0.125		
e	1.1250	2	0.625		
e^{Δ}	1.375	3	0.458		

由于 $F_{0.05}(1,3)=10.13$, $F_{0.01}(1,3)=34.12$,故因素 A，B 作用高度显著，因素 C 作用不显著，因素 D 作用显著，这与前面极差分析的结果是一致的.F 检验法要求选取 S_e，且希望 f_e 要大，故在安排试验时，适当留出些空列会有好处的.前面的方差分析中，讨论因素 A 和 B 的交互作用 $A\times B$.这类交互作用在正交试验设计中同样有表现，即一个因素 A 的水平对试验结果指标的影响同另一个因素 B 的水平选取有关.当试验考虑交互作用时，也可用前面讲的基本方法来处理.本章就不再介绍了.

小　结

本章介绍了数理统计的基本方法之一：方差分析.

在生产实践中，试验结果往往要受到一种或多种因素的影响.方差分析就是通过对试验数据进行分析，检验方差相同的多个正态总体的均值是否相等，用以判断各因素对试验结果的影响是否显著.方差分析按影响试验结果的因素的个数分为单因素方差分析、双因素方差分析和多因素方差分析.

1. 单因素方差分析的情况.试验数据总是参差不齐，我们用总偏差平方和 $S_T=\sum_{j=1}^{s}\sum_{i=1}^{n_j}(x_{ij}-\overline{x})^2$ 来度量数据间的离散程度.将 S_T 分解为试验随机误差的平方和(S_E)与因素 A 的偏差平方和（S_A）之和.若 S_A 比 S_E 大得较多，则有理由认为因素的各个水平对应的试验结果有显著差异，从而拒绝因素各水平对应的正态总体的均值相等这一原假设.这就是单因素方差分析法的基本思想.

2. 双因素方差分析的基本思想类似于单因素方差分析.但双因素试验的方差分析中，我们不仅要检验因素 A 和 B 各自的作用，还要检验它们之间的交互作用.

3. 正交试验设计及其方差分析.根据因素的个数及各个因素的水平个数，选取适当的正交表并按表进行试验.我们通过对这少数的试验数据进行分析，推断出各因素对试验结果影响的大小.对正交试验结果的分析，通常采用两种方法，一种是直观分析法（极差分析法），它通过对各因素极差 R_j 的排序来确定各因素对试验结果影响的大小.一种是方差分析法，它的基本思想类似于双因素的方差分析.

重要术语及主题

单因素试验方差分析的数学模型　　　　$S_T=S_E+S_A$

单因素方差分析表　　双因素方差分析表　　正交试验表　　极差分析表

习 题 九

1. 灯泡厂用 4 种不同的材料制成灯丝，检验灯线材料这一因素对灯泡寿命的影响.若灯泡寿命服从正态分布，不同材料的灯丝制成的灯泡，其寿命的方差相同，试根据表 9-26 中的试验结果记录，在显著性水平 0.05 下检验灯泡寿命是否因灯丝材料不同而有显著差异？

表 9-26

		试验批号							
		1	2	3	4	5	6	7	8
灯丝	A_1	1600	1610	1650	1680	1700	1720	1800	
材料	A_2	1580	1640	1640	1700	1750			
水平	A_3	1460	1550	1600	1620	1640	1660	1740	1820
	A_4	1510	1520	1530	1570	1600	1680		

2. 一个年级有三个小班，他们进行了一次数学考试，现从各个班级随机地抽取了一些学生，记录其成绩如表 9-27 所示：

表 9-27

I	II	III
73　66	88　77	68　41
89　60	78　31	79　59
82　45	48　78	56　68
43　93	91　62	91　53
80　36	51　76	71　79
73　77	85　96	71　15
	74　80	87
	56	

试在显著性水平 0. 05 下检验各班级的平均分数有无显著差异.设各个总体服从正态分布，且方差相等.

3.表 9-28 记录了 3 位操作工分别在不同机器上操作 3 天的日产量.

表 9-28

操作工 / 机器	甲	乙	丙
A_1	15　15　17	19　19　16	16　18　21
A_2	17　17　17	15　15　15	19　22　22
A_3	15　17　16	18　17　16	18　18　18
A_4	18　20　22	15　16　17	17　17　17

取显著性水平 α=0.05,试分析操作工之间，机器之间以及两者交互作用有无显著差异？

4. 为了解 3 种不同配比的饲料对仔猪生长影响的差异，对 3 种不同品种的猪各选 3 头进行试验，分别测得其 3 个月间体重增加量如表 9-29 所示，取显著性水平 α=0.05，试分析不同饲料与不同品种对猪的生长有无显著影响？假定其体重增长量服从正态分布，且各种配比的方差相等.

表 9-29

体重增长量		因素 B（品种）		
		B_1	B_2	B_3
因	A_1	51	56	45
素	A_2	53	57	49
A（饲料）	A_3	52	58	47

5. 研究氯乙醇胶在各种硫化系统下的性能（油体膨胀绝对值越小越好）需要考察补强剂（A）、防老剂（B）、硫化系统（C）3 个因素（各取 3 个水平），根据专业理论经验，交互作用全忽略，根据选用 $L_9(3^4)$表作 9 次试验及试验结果见表 9-30：

表 9-30

表头设计					试验
列号 试验号	1	2	3	4	结果
1	1	1	1	1	7.25
2	1	2	2	2	5.48
3	1	3	3	3	5.35
4	2	1	2	3	5.40
5	2	2	3	1	4.42
6	2	3	1	2	5.90
7	3	1	3	2	4.68
8	3	2	1	3	5.90
9	3	3	2	1	5.63

(1) 试作最优生产条件的直观分析，并对 3 因素排出主次关系.

(2) 给定 α=0.05,作方差分析与(1)比较.

6. 某农科站进行早稻品种试验（产量越高越好），需考察品种（A），施氮肥量（B），氮、磷、钾肥比例（C），插植规格（D）4 个因素，根据专业理论和经验，交互作用全忽略，早稻试验方案及结果分析见表 9-31：

表 9-31

因素试验号	A 品种	B 施氮肥量	C 氮、磷、钾肥比例	D 插植规格	试验指标 产量
1	1（科 6 号）	1(20)	1(2∶2∶1)	1(5×6)	19.0
2	1	2(25)	2(3∶2∶3)	2(6×6)	20.0
3	2(科 5 号)	1	1	2	21.9
4	2	2	2	1	22.3
5	1(科 7 号)	1	2	1	21.0
6	1	2	1	2	21.0
7	2(珍珠矮)	1	2	2	18.0
8	2	2	1	1	18.2

(1) 试作出最优生产条件的直观分析，并对 4 因素排出主次关系.

(2) 给定 α=0.05,作方差分析，与(1)比较.