安徽理工大學
ANHUI UNIVERSITY OF SCIENCE & TECHNOLOGY

研究生课程

试验设计与分析

主讲 闵凡飞 教授

材料科学与工程学院

2020年5月

第四章 试验结果的分析
-方差分析

4.2 试验结果的方差分析

1. 方差分析的基本原理

由英国统计学家R.A.Fisher首创，为纪念Fisher以F命名，故方差分析又称 F 检验 （F test）。

用于推断多个总体均值有无差异。

方差分析法是将因素水平（或交互作用）的变化所引起的试验结果间的差异与误差波动所引起的试验结果间的差异区分开来的一种数学方法 。

1、单因素方差分析原理

单因素方差分析是固定其它因素只考虑某一因素A对试验指标的影响。为此将因素A以外的条件保持不变，取因素A的r个水平A_1，A_2，…，A_r，对水平A*i*（*i*=1，2，…，r）重复做n_i次试验，得到试验指标的观察值列于下表。

单因素方差分析试验指标观察值

水平	1	2	…	
A_1	x_{11}	x_{12}	…	x_{1n_1}
A_2	x_{21}	x_{22}	…	x_{2n_2}
⋮	⋮	⋮	…	⋮
A_i	x_{i1}	x_{i2}	…	x_{in_i}
⋮	⋮	⋮	…	⋮
A_r	x_{r1}	x_{r2}	…	x_{rn_r}

我们假定在各个水平 A_i（i=1，2，…，r）下的样本为 X_{i1}，X_{i2}，…，X_{in_i}，它们来自具有相同方差 σ^2，均值分别为 μ_i 的正态总体 $X_i \sim N(\mu_i, \sigma^2)$，其中 μ_i，σ^2 均为未知，并且不同水平 A_i 下的样本之间相互独立。

我们取下面的线性统计模型

$$\left.\begin{aligned} x_{ij} &= \mu_i + \varepsilon_{ij} \\ \varepsilon_{ij} &\sim N(0, \sigma^2) \end{aligned}\right\},$$

其中 ε_{ij} 为随机误差；i=1，2，…，r；j=1，2，…，n_i

设

$$\mu = \frac{1}{n}\sum_{i=1}^{r} n_i \mu_i$$

为总平均值，其中 $n = \sum_{i=1}^{r} n_i$。

令

$$\delta_i = \mu_i - \mu$$

为第 i 个水平 A_i 的效应，$\sum_{i=1}^{r} n_i \delta_i = 0$，则有

$$\left.\begin{aligned} x_{ij} &= \mu + \delta_i + \varepsilon_{ij} \\ \varepsilon_{ij} &\sim N(0, \sigma^2) \end{aligned}\right\},$$

其中 i=1，2，…，r；j=1，2，…，n_i。

方差分析的任务就是检验线性统计模型中 r 个总体 $N(\mu_i, \sigma^2)$ 中的各 μ_i 的相等性,即有

原假设 H_0：$\mu_1 = \mu_2 = \cdots = \mu_r$，

对立假设 H_1：$\mu_i \neq \mu_j$ 至少有一对这样的 i，j，

也就是下面的等价假设：

$$H_0: \ \delta_1 = \delta_2 = \cdots = \delta_r = 0$$

H_1：$\delta_i \neq 0$ 至少有一个 i，

检验这种假设的适当程序就是方差分析。

具体过程如下：

1、总离差平方和的分解

记水平 A_i 下的样本均值（试验结果平均值）为

$$\bar{x}_i=\frac{1}{n_i}\sum_{j=1}^{n_i}x_{ij}$$

样本数据的总平均值为

$$\bar{x}=\frac{1}{n}\sum_{i=1}^{r}\sum_{j=1}^{n_i}x_{ij}$$

总离差平方和为

$$S_T=\sum_{i=1}^{r}\sum_{j=1}^{n_i}(x_{ij}-\bar{x})^2$$

将 S_T 改写并分解得

$$S_T=\sum_{i=1}^{r}\sum_{j=1}^{n_i}[(\bar{x}_i-\bar{x})+(x_{ij}-\bar{x}_i)]^2$$

$$=\sum_{i=1}^{r}\sum_{j=1}^{n_i}(\bar{x}_i-\bar{x})^2+\sum_{i=1}^{r}\sum_{j=1}^{n_i}(x_{ij}-\bar{x}_i)^2+2\sum_{i=1}^{r}\sum_{j=1}^{n_i}(\bar{x}_i-\bar{x})(x_{ij}-\bar{x}_i)$$

上面展开式中的第三项为 0。

因为

$$2\sum_{i=1}^{r}\sum_{j=1}^{n_i}(\bar{x}_i-\bar{x})(x_{ij}-\bar{x}_i)=2\sum_{i=1}^{r}(\bar{x}_i-\bar{x})\sum_{j=1}^{n_i}(x_{ij}-\bar{x}_i)$$

$$=2\sum_{i=1}^{r}(\bar{x}_i-\bar{x})(\sum_{j=1}^{n_i}x_{ij}-n_i\bar{x}_i)=0$$

若记

$$S_A=\sum_{i=1}^{r}\sum_{j=1}^{n_i}(\bar{x}_i-\bar{x})^2$$

$$S_e=\sum_{i=1}^{r}\sum_{j=1}^{n_i}(x_{ij}-\bar{x}_i)^2$$

则有

$$S_T=S_A+S_e$$

这里 S_T表示全部试验数据与总平均值之间的差异，又叫总变差平方和。S_A表示在 A_i水平下的试验数据平均值与总平均值之间的差异，叫因素 A 效应平方和，又叫组间差，它主要是由试验条件改变引起的。S_e表示在 A_i水平下试验数据与该水平下试验数据平均值间的差异，它是由随机误差引起的，叫误差平方和，又叫组内差。

2、统计分析

由于

$$x_{ij} \sim N(\mu_i, \sigma^2)$$

将 S_T 改写为下面的形式

$$S_T = \sum_{i=1}^{r}\sum_{j=1}^{n_i}(x_{ij}-\bar{x})^2 = (n-1)S^2$$

这里 S^2 是样本方差，即

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{r}\sum_{j=1}^{n_i}(x_{ij}-\bar{x})^2$$

考虑到

$$\frac{S_T}{\sigma^2} = \frac{(n-1)S^2}{\sigma^2} \sim \chi^2(n-1),$$

从这里知道 S_T 的自由度为（n-1）。

将 S_e 改写为下面的形式

$$S_e = \sum_{i=1}^{r}\sum_{j=1}^{n_i}(x_{ij}-\bar{x}_i)^2 = \sum_{i=1}^{r}(n_i-1)S_i^2$$

这里 S_i^2 是在 A_i 水平下的样本方差，即

$$S_i^2 = \frac{1}{n_i-1}\sum_{j=1}^{n_i}(x_{ij}-\bar{x}_i)^2$$

因为

$$\frac{(n_i-1)S_i^2}{\sigma^2} \sim \chi^2(n_i-1)$$

再由 χ^2 分布的可加性知

$$\frac{S_e}{\sigma^2}=\sum_{i=1}^{r}\frac{(n_i-1)S_i^2}{\sigma^2}\sim\chi^2[\sum_{i=1}^{r}(n_i-1)]$$

即

$$\frac{S_E}{\sigma^2}\sim\chi^2(n-r)$$

由此可知，S_e的自由度为（n-r），并且有

$$E\left(\frac{S_E}{\sigma^2}\right)=n-r\text{，}$$

即有

$$E(S_E)=(n-r)\sigma^2$$

$$S_A=\sum_{i=1}^{r}\sum_{j=1}^{n_i}(\bar{x}_i-\bar{x})^2=\sum_{i=1}^{r}n_i(\bar{x}_i-\bar{x})^2\text{，}$$

展开后可化成

$$S_A=\sum_{i=1}^{r}n_i\bar{x}_i^2-n\bar{x}^2$$

由相关式和 x_{ij} 之间的独立性可知

$$\bar{x}_i\sim N(\mu_i,\frac{\sigma^2}{n_i})\text{，}$$

D(x)=E(x²)-E²(x)

$$\bar{x} \sim N(\mu, \frac{\sigma^2}{n}),$$

所以 $E(\bar{x}_i) = \mu_i$，$V(\bar{x}_i) = \frac{\sigma^2}{n_i}$，$E(\bar{x}) = \mu$，$V(\bar{x}) = \frac{\sigma^2}{n}$。

再由 $E(\bar{x}_i^2) = V(\bar{x}_i) + E^2(\bar{x}_i)$，$E(\bar{x}^2) = V(\bar{x}) + E^2(\bar{x})$，得

$$E(S_A) = E[\sum_{i=1}^{r} n_i \bar{x}_i^2 - n\bar{x}^2] = \sum_{i=1}^{r} n_i E(\bar{x}_i^2) - nE(\bar{x}^2)$$

$$= \sum_{i=1}^{r} n_i \left[\frac{\sigma^2}{n_i} + \mu_i^2\right] - n\left[\frac{\sigma^2}{n} + \mu^2\right]$$

$$= r\sigma^2 + \sum_{i=1}^{r} n_i(\mu + \delta_i)^2 - \sigma^2 - n\mu^2$$

$$= (r-1)\sigma^2 + \sum_{i=1}^{r} n_i \mu^2 + 2\mu \sum_{i=1}^{r} n_i \delta_i + \sum_{i=1}^{r} n_i \delta_i^2 - n\mu^2 。$$

由于 $\sum_{i=1}^{r} n_i = n$，$\sum_{i=1}^{r} n_i \delta_i = 0$，所以得出

$$E(S_A) = (r-1)\sigma^2 + \sum_{i=1}^{r} n_i \delta_i^2$$

在 H_0：$\delta_i = 0$ 成立的条件下，

$$E(S_A) = (r-1)\sigma^2$$

$$E\left(\frac{S_A}{r-1}\right)=\sigma^2$$

因为 S_A 与 S_e 相互独立（证明略），再由 χ^2 分布的可加性性质可得出

$$\frac{S_A}{\sigma^2}\sim\chi^2(r-1)$$

并得出 S_A 的自由度为（r-1）。

记

$$V_A=\frac{S_A}{r-1},$$

$$V_e=\frac{S_e}{n-r},$$

并分别叫做 S_A，S_e 的均方。由此可知，V_e 是 σ^2 的无偏估计，当 H_0 成立时， V_A 也是 σ^2 的无偏估计。

在 H_0 成立的条件下，取统计量

$$F=\frac{\frac{S_A}{\sigma^2}\Big/ r-1}{\frac{S_e}{\sigma^2}\Big/ n-r}\sim F(r-1,n-r),$$

即

$$F=\frac{V_A}{V_e}\sim F(r-1,n-r)$$

对于给出的 α，从 F 分布表中查出 $F_\alpha(r-1,n-r)$的值，由样本值计算出 S_A，S_e，从而算出 F 值。若 H_0不成立，即$\delta_i \neq 0$（至少一个 i），S_A偏大，导致 F 偏大，因此，判断如下：

若 $F>F_\alpha(r-1,n-r)$，则拒绝 H_0；若 $F<F_\alpha(r-1,n-r)$，则接受 H_0。

将上面的分析过程和结果，列成一个简洁的表格（表 4-29），可以对整个方差分析过程进行很好的汇总，这个表叫做方差分析表。

表 4-29 单因素方差分析表

方差来源	离差平方和	自由度	均方	F比	显著性
因素 A	S_A	$r-1$	$V_A=\dfrac{S_A}{r-1}$	$F=\dfrac{V_A}{V_e}$	
误差 E	S_E	$n-r$	$V_e=\dfrac{S_e}{n-r}$		
总和 T	S_T	$n-1$			

2、双因素有交互作用方差分析原理

设两因素A，B，A有r个水平，A_1，A_2，…，A_r，B有k个水平：B_1，B_2，…，B_k，在每一个组合水平（A_i，B_j）下重复做n次（$n \geq 2$）试验，每个观察值记为x_{ijm}，结果如下表 。

双因素有交互作用方差分析试验指标观察值

B_j \ A_i	B_1	B_2	…	B_k
A_1	x_{111}　x_{112}　…　x_{11n}	x_{121}　x_{122}　…　x_{12n}	…	x_{1k1}　x_{1k2}　…　x_{1kn}
A_2	x_{211}　x_{212}　…　x_{21n}	X_{221}　x_{222}　…　x_{22n}	…	X_{2k1}　x_{2k2}　…　x_{2kn}
⋮	⋮	⋮	…	⋮
A_r	x_{r11}　x_{r12}　…　x_{r1n}	X_{r21}　x_{r22}　…　x_{r2n}	…	X_{rk1}　x_{rk2}　…　x_{rkn}

设 $x_{ijm} \sim N(\mu_{ij}, \sigma^2)$，各 x_{ij} 之间相互独立，i=1，2，…，r，j=1，2，…，k，m=1，2，…，n，各 x_{ijm} 相互独立。

并设

$$\mu_{ij} = \mu + \alpha_i + \beta_j + \gamma_{ij}$$

其中 μ 为总平均值，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，显然有

$$\sum_{i=1}^{r}\alpha_i = 0，\sum_{j=1}^{k}\beta_j = 0，\sum_{i=1}^{r}\gamma_{ij} = 0，\sum_{j=1}^{k}\gamma_{ij} = 0$$

我们取下面的线性统计模型

$$\left.\begin{array}{l} x_{ijm} = \mu + \alpha_i + \beta_j + \gamma_{ij} + \varepsilon_{ijm} \\ \varepsilon_{ijm} \sim N(0, \sigma^2) \\ \sum\limits_{i=1}^{r}\alpha_i = 0, \sum\limits_{j=1}^{k}\beta_j = 0, \sum\limits_{i=1}^{r}\gamma_{ij} = 0 \sum\limits_{j=1}^{k}\gamma_{ij} = 0 \end{array}\right\}，$$

其中各 ε_{ijm} 相互独立；i=1，2，…，r；j=1，2，…，k，m=1，2，…，n；μ，α_i，β_j，γ_{ij}，σ^2 都是未知参数。

对这个线性模型，我们检验如下假设

$$\left.\begin{array}{l} H_{A0}: \alpha_1 = \alpha_2 = \cdots = \alpha_r = 0 \\ H_{A1}: \alpha_i \neq 0,\ \text{至少一个} i \end{array}\right\},$$

$$\left.\begin{array}{l} H_{B0}: \beta_1 = \beta_2 = \cdots = \beta_k = 0 \\ H_{B1}: \beta_j \neq 0,\ \text{至少一个} j \end{array}\right\},$$

$$\left.\begin{array}{l} H_{AB0}: \gamma_{ij} = 0, i = 1,2,\cdots,r, j = 1,2,\cdots,k \\ H_{AB1}: \gamma_{ij} \neq 0,\ \text{至少一对} i, j \end{array}\right\},$$

1、总离差平方和的分解

记

$$\bar{x}_i = \frac{1}{kn}\sum_{j=1}^{k}\sum_{m=1}^{n} x_{ijm}$$

$$\bar{x}_j=\frac{1}{rn}\sum_{i=1}^{r}\sum_{m=1}^{n}x_{ijm}$$

$$\bar{x}_{ij}=\frac{1}{n}\sum_{m=1}^{n}x_{ijm}$$

$$\bar{x}=\frac{1}{rkn}\sum_{i=1}^{r}\sum_{j=1}^{k}\sum_{m=1}^{n}x_{ijm}$$

总离差平方和为

$$S_T=\sum_{i=1}^{r}\sum_{j=1}^{k}\sum_{m=1}^{n}(x_{ijm}-\bar{x})^2$$

将 S_T改写并分解得

$$S_T=\sum_{i=1}^{r}\sum_{j=1}^{k}\sum_{m=1}^{n}[(\bar{x}_i-\bar{x})+(\bar{x}_j-\bar{x})+(\bar{x}_{ij}-\bar{x}_i-\bar{x}_j+\bar{x})+(x_{ijm}-\bar{x}_{ij})]^2$$

$$=kn\sum_{i=1}^{r}(\bar{x}_i-\bar{x})^2+rn\sum_{j=1}^{k}(\bar{x}_j-\bar{x})^2+n\sum_{i=1}^{r}\sum_{j=1}^{k}(\bar{x}_{ij}-\bar{x}_i-\bar{x}_j+\bar{x})^2+\sum_{i=1}^{r}\sum_{j=1}^{k}\sum_{m=1}^{n}(x_{ijm}-\bar{x}_{ij})^2$$

若记

$$S_A = kn\sum_{i=1}^{r}(\bar{x}_i - \bar{x})^2$$

$$S_B = rn\sum_{j=1}^{k}(\bar{x}_j - \bar{x})^2$$

$$S_{A\times B} = n\sum_{i=1}^{r}\sum_{j=1}^{k}(\bar{x}_{ij} - \bar{x}_i - \bar{x}_j + \bar{x})^2$$

$$S_E = \sum_{i=1}^{r}\sum_{j=1}^{k}\sum_{m=1}^{n}(x_{ijm} - \bar{x}_{ij})^2$$

则有

$$S_T = S_A + S_B + S_{A\times B} + S_E$$

S_A，S_B分别为因素 A、因素 B 效应的平方和，$S_{A\times B}$为因素 A，B 的交互效应平方和，S_E为误差平方和。

2、统计分析

首先，这里，S_T的自由度为（rkn-1），S_A的自由度为（r-1），S_B的自由度为（k-1），$S_{A\times B}$的自由度为$(r-1)(k-1)$，S_E的自由度为 $rk(n-1)$。

相应地有均方值

$$\bar{S}_A = \frac{S_A}{r-1},$$

$$\bar{S}_B = \frac{S_B}{k-1},$$

$$\bar{S}_{A\times B} = \frac{S_{A\times B}}{(r-1)(k-1)},$$

$$\bar{S}_E = \frac{S_E}{rk(n-1)},$$

它们的期望值分别为

$$E(\bar{S}_A) = \sigma^2 + \frac{kn}{r-1}\sum_{i=1}^{r}\alpha_i^2$$

$$E(\bar{S}_B) = \sigma^2 + \frac{rn}{k-1}\sum_{j=1}^{k}\beta_j^2$$

$$E(\bar{S}_{A\times B}) = \sigma^2 + \frac{n}{(r-1)(k-1)}\sum_{i=1}^{r}\sum_{j=1}^{k}\gamma_{ij}^2$$

$$E(\bar{S}_E) = \sigma^2$$

当原假设 H_0 都成立时，$\bar{S}_A$，$\bar{S}_B$，$\bar{S}_{A\times B}$，$\bar{S}_E$ 都是 σ^2 的无偏估计量。

在 H_{A0} 成立的条件下，取统计量

$$F_A = \frac{\bar{S}_A}{\bar{S}_E} \sim F(r-1, rk(n-1))$$

在 H_{B0} 成立的条件下，取统计量

$$F_B = \frac{\bar{S}_B}{\bar{S}_E} \sim F(k-1, rk(n-1))$$

在 H_{AB0} 成立的条件下，取统计量

$$F_{A\times B} = \frac{\bar{S}_{A\times B}}{\bar{S}_E} \sim F((r-1)(k-1), rk(n-1))$$

由试验数据可以计算出 F_A，F_B 和 $F_{A\times B}$ 的值。

对于给出的 α，从 F 分布表中查出 $F_\alpha(r\text{-}1, rk(n\text{-}1))$，$F_\alpha(k\text{-}1, rk(n\text{-}1))$ 和 $F_\alpha((r\text{-}1)(k\text{-}1), rk(n\text{-}1))$

的值，若 $F_A > F_\alpha(r-1, rk(n-1))$，则拒绝 H_{A0}，否则，就接受 H_{A0}；若 $F_B > F_\alpha(k-1, rk(n-1))$，则拒绝 H_{B0}，否则，就接受 H_{B0}；若 $F_{A\times B} > F_\alpha((r-1)(k-1), rk(n-1))$，则拒绝 H_{AB0}，否则，就接受 H_{AB0}

为了计算的方便，通常采用下面的简便计算公式：

记 $T_i = \sum_{j=1}^{k}\sum_{m=1}^{n} x_{ijm}$ $i=1,2,\cdots,r$，$T_j = \sum_{i=1}^{r}\sum_{m=1}^{n} x_{ijm}$ $j=1,2,\cdots,k$，$T = \sum_{i=1}^{r}\sum_{j=1}^{k}\sum_{m=1}^{n} x_{ijm}$，

$T_{ij} = \sum_{m=1}^{n} x_{ijm}$，$Q = \sum_{i=1}^{r}\sum_{j=1}^{k}\sum_{m=1}^{n} x_{ijm}^2$，

则有

$$S_T = Q - \frac{T^2}{rkn}$$

$$S_A = \frac{1}{kn}\sum_{i=1}^{r} T_i^2 - \frac{T^2}{rkn}$$

$$S_B = \frac{1}{rn}\sum_{j=1}^{k} T_j^2 - \frac{T^2}{rkn}$$

$$S_{A\times B} = \frac{1}{n}\sum_{i=1}^{r}\sum_{j=1}^{k} T_{ij}^2 - \frac{T^2}{rkn} - S_A - S_B$$

双因素有交互作用方差分析表见表 3.2.6。

表 3.2.6 双因素有交互作用方差分析表

方差来源	离差平方和	自由度	均方	F 比	显著性
因素 A	S_A	$r-1$	$\bar{S}_A=\dfrac{S_A}{r-1}$	$F_A=\dfrac{\bar{S}_A}{\bar{S}_E}$	
因素 B	S_B	$k-1$	$\bar{S}_B=\dfrac{S_B}{k-1}$	$F_B=\dfrac{\bar{S}_B}{\bar{S}_E}$	
因素 $A\times B$	$S_{A\times B}$	$(r-1)(k-1)$	$\bar{S}_{A\times B}=\dfrac{S_{A\times B}}{(r-1)(k-1)}$	$F_{A\times B}=\dfrac{\bar{S}_{A\times B}}{\bar{S}_E}$	
误差 E	S_E	$rk(n-1)$	$\bar{S}_E=\dfrac{S_E}{rk(n-1)}$		
总和 T	S_T	$rkn-1$			

2. 正交试验结果方差分析的步骤与格式

设用正交表安排m个因素的试验，试验次数为n，试验结果分别为$x_1,x_2,\ldots,x_n$。假定每个因素有n_a个水平，每个水平做a次试验，则$n=an_a$；则正交试验设计的方差分析一般步骤和格式如下：

1、计算总偏差平方和

$$S_T=\sum_{k=1}^{n}(x_k-\bar{x})^2=\sum_{k=1}^{n}x_k^2-n\bar{x}^2=Q-\frac{T^2}{n}$$

其中

$\bar{x}=\frac{1}{n}\sum_{k=1}^{n}x_k$ 为所有试验结果的总平均值，$Q=\sum_{k=1}^{n}x_k^2$ 为所有试验结果平方的和，

$T=\sum_{k=1}^{n}x_k$ 为所有试验结果的和。

下面以计算因素 A 的离差平方和为例来说明如何计算各因素的离差平方和。

设因素 A 安排在正交表的第 i 列，可看作单因素试验，用 x_{ij}表示因素（正交表）第 i 列 j 水平对应的试验结果的总和。则：

$$S_A = S_i = \frac{1}{a}\sum_{j=1}^{p} x_{ij}^2 - n\bar{x}^2 = Q_A - \frac{T^2}{n}$$

其中

S_i　表示正交表第 i 列的离差平方和；

a　表示第 i 列同水平试验次数；

p　表示第 i 列的水平数；

x_{ij}　表示因素（正交表）第 i 列 j 水平（i=1, 2, …, n$_i$；对应的试验结果的总和（可以在正交表中计算出来）

$$Q_A = \frac{1}{a}\sum_{j=1}^{p} x_{ij}^2$$

S_A反映了因素A水平变化时所引起的试验结果的差异，即因素A对试验指标的影响。相同的方法可以分别计算出正交表中其余各列的离差平方和。各因素的偏差平方和与所在正交表相应列的偏差平方和相等。对于因素间的交互作用，如果占两列或以上时，则交互作用的偏差平方和等于所占列偏差平方和之和。比如因素 A 和 B 的交互作用占正交表的 3、4 两列，则

$$S_{A\times B}=S_3+S_4$$

3、计算试验误差的平方和 S_E

设$S_{因+交}$为所有因素及要考虑的交互作用的偏差平方和，因为

$$S_T=S_{因+交}+S_E$$

所以

$$S_E=S_T-S_{因+交}$$

4、计算自由度

根据自由度的概念，各自由度可按下面公式计算

1）总自由度 f_T=总的试验次数－1＝n-1

2）因素的自由度 $f_{因}$=因素水平数－1＝n_a-1

3）因素交互作用的自由度等于因素的自由度之积，例如

$$f_{A\times B} = f_A \times f_B$$

4）试验误差的自由度 f_E=f_T−$f_{因+交}$

5、计算平均离差平方和

因素平均离差平方和

$$\bar{S}_{因} = \frac{S_{因}}{f_{因}}$$

交互作用平均离差平方和

$$\bar{S}_{交} = \frac{S_{交}}{f_{交}}$$

试验误差平均平方和

$$\bar{S}_E = \frac{S_E}{f_E}$$

6、求统计量$F_{比}$

$$F_{比}=\frac{\overline{S}_{因或交}}{\overline{S}_{E}}$$

7、显著性检验

1）若 $F_{因或交}>F_{0.01}(f_{因或交},f_E)$ 则说明因素或交互作用水平的改变对试验指标有高度显著的影响，记作:×××;

2）若 $F_{0.05}(f_{因或交},f_E)<F_{因或交}<F_{0.01}(f_{因或交},f_E)$ 则说明因素或交互作用水平的改变对试验指标有显著的影响，记作：××;

3）若 $F_{0.1}(f_{因或交},f_E)<F_{因或交}<F_{0.05}(f_{因或交},f_E)$ 则说明因素或交互作用水平的改变对试验指标有一定影响，记作：×

8、列出方差分析表

正交试验设计方差分析表

方差来源	离差平方和	自由度	均方	F 比	显著性
因素	$S_{因}$	n_a-1	$\bar{S}_{因}=\dfrac{S_{因}}{n_a-1}$	$F_{因}=\dfrac{\bar{S}_{因}}{\bar{S}_E}$	
交互作用	$S_{交}$	交互作用因素自由度之积	$\bar{S}_{交}=\dfrac{S_{交}}{f_{交}-1}$	$F_{交}=\dfrac{\bar{S}_{交}}{\bar{S}_E}$	
误差 E	S_E	$f_T-f_{因}-f_{交}$	$\bar{S}_E=\dfrac{S_E}{n-r}$		
总和 T	S_T	$n-1$			

3. 三水平含有交互作用的正交试验结果的方差分析

实例：为了提高某产品的产量，需要考察3个因素：反应温度、反应压力和溶液浓度，每个因素都取3个水平，具体数值如下表。同时考察因素间所有的一级交互作用，试进行方差分析确定所考察因素对试验指标产品产量的影响规律。

因素及水平表

因素水平	*A* 温度 ℃	*B* 压力 10^5Pa	*D* 浓度 %
1	60	2.0	0.5
2	65	2.5	1.0
3	70	3.0	2.0

试验结果分析

正交试验安排及试验结果

表头设计	A	B	A×B		C	A×C		B×C 1			B×C 2			试验结果
试验号 \ 列号	1	2	3	4	5	6	7	8	9	10	11	12	13	
1	1	1	1	1	1	1	1	1	1	1	1	1	1	1.30
2	1	1	1	1	2	2	2	2	2	2	2	2	2	4.65
3	1	1	1	1	3	3	3	3	3	3	3	3	3	7.23
4	1	2	2	2	1	1	1	2	2	2	3	3	3	0.50
5	1	2	2	2	2	2	2	3	3	3	1	1	1	3.67
6	1	2	2	2	3	3	3	1	1	1	2	2	2	6.23
7	1	3	3	3	1	1	1	3	3	3	2	2	2	1.37
8	1	3	3	3	2	2	2	1	1	1	3	3	3	4.73
9	1	3	3	3	3	3	3	2	2	2	1	1	1	7.07
10	2	1	2	3	1	2	3	1	2	3	1	2	3	0.47
11	2	1	2	3	2	3	1	2	3	1	2	3	1	3.47
12	2	1	2	3	3	1	2	3	1	2	3	1	2	6.13
13	2	2	3	1	1	2	3	2	3	1	3	1	2	0.33
14	2	2	3	1	2	3	1	3	1	2	1	2	3	3.40
15	2	2	3	1	3	1	2	1	2	3	2	3	1	5.80
16	2	3	1	2	1	2	3	3	1	2	2	3	1	0.63
17	2	3	1	2	2	3	1	1	2	3	3	1	2	3.97
18	2	3	1	2	3	1	2	2	3	1	1	2	3	6.50
19	3	1	3	2	1	3	2	1	3	2	1	3	2	0.03
20	3	1	3	2	2	1	3	2	1	3	2	1	3	3.40
21	3	1	3	2	3	2	1	3	2	1	3	2	1	6.80
22	3	2	1	3	1	3	2	2	1	3	3	2	1	0.57
23	3	2	1	3	2	1	3	3	2	1	1	3	2	3.97
24	3	2	1	3	3	2	1	1	3	2	2	1	3	6.83
25	3	3	2	1	1	3	2	3	2	1	2	1	3	1.07
26	3	3	2	1	2	1	3	1	3	2	3	2	1	3.97
27	3	3	2	1	3	2	1	2	1	3	1	3	2	6.57
K_1	36.73	33.46	35.63	34.30	6.27	32.94	34.21	33.33	32.96	34.40	32.98	33.77	33.28	
K_2	30.70	31.30	32.08	31.73	35.21	34.66	33.13	33.04	34.30	33.21	33.43	33.96	33.25	
K_3	33.21	35.88	32.93	34.61	59.16	33.04	33.30	34.27	33.38	33.03	34.23	32.91	34.11	
$\overline{K}_1$	4.08	3.72	3.96	3.81	0.70	3.66	3.80	3.70	3.66	3.82	3.66	3.75	3.70	
$\overline{K}_2$	3.41	3.48	3.56	3.53	3.91	3.85	3.68	3.67	3.81	3.69	3.71	3.77	3.69	
$\overline{K}_3$	3.69	3.99	3.66	3.85	6.57	3.67	3.70	3.81	3.71	3.67	3.80	3.66	3.79	
R	0.67	0.51	0.40	0.32	5.87	0.19	0.12	0.14	0.15	0.15	0.14	0.11	0.10	
S_j	2.04	1.17	0.76	0.56	155.87	0.21	0.08	0.09	0.10	0.12	0.09	0.07	0.05	

试验结果的方差分析

1、计算离差平方和

1）计算总离差平方和

$$S_T=\sum_{k=1}^{n}(x_k-\bar{x})^2=\sum_{k=1}^{n}x_k^2-n\bar{x}^2=\sum_{k=1}^{27}x_k^2-27\bar{x}^2=161.20$$

$$\bar{x}=\frac{1}{n}\sum_{k=1}^{n}x_k=\frac{1}{27}\sum_{k=1}^{27}x_k=3.73$$

2）计算因素的离差平方和

$$S_i=\frac{1}{a}\sum_{j=1}^{p}x_{ij}^{2}-n\bar{x}^2=\frac{1}{9}\sum_{j=1}^{3}x_{ij}^2-27\bar{x}^2$$

由此可以分别计算出：

$$S_A=S_1=2.04$$
$$S_B=S_2=1.17$$
$$S_{A\times B}=S_3+S_4=1.32$$
$$S_C=S_5=155.87$$
$$S_{A\times C}=S_6+S_7=0.28$$
$$S_{C\times B}=S_8+S_{11}=0.18$$
$$S_9=0.10$$
$$S_{10}=0.12$$
$$S_{12}=0.07$$
$$S_{13}=0.05$$

（3）计算试验误差的平方和

$$Se=S_T-S_A-S_B-S_C-S_{A\times B}-S_{A\times C}-S_{B\times C}=S_9+S_{10}+S_{12}+S_{13}=0.34$$

2、计算自由度

1） f_T=总的试验次数－1＝27-1＝26

2） $f_A=f_B=f_C$＝水平数－1＝3-1＝2

3） $f_{A\times B}=f_{A\times C}=f_{C\times B}=2\times2=4$

4） $f_E=f_T-f_A-f_B-f_C-f_{A\times B}-f_{A\times C}-f_{B\times C}=8$

3、计算平均离差平方和

$$\bar{S}_A=\frac{S_A}{f_A}=1.02$$

$$\bar{S}_B=\frac{S_B}{f_B}=0.58$$

$$\bar{S}_C=\frac{S_C}{f_C}=77.93$$

$$\bar{S}_{A\times B}=\frac{S_{A\times B}}{f_{A\times B}}=0.33$$

$$\bar{S}_{A\times C}=\frac{S_{A\times C}}{f_{A\times C}}=0.07$$

$$\bar{S}_{B\times C}=\frac{S_{B\times C}}{f_{B\times C}}=0.05$$

$$\bar{S}_{E}=\frac{S_{E}}{f_{A}}=0.04$$

4）求$F_{比}$

$$F_{A}=\frac{\bar{S}_{A}}{\bar{S}_{E}}=25.50$$

$$F_{B}=\frac{\bar{S}_{B}}{\bar{S}_{E}}=14.50$$

$$F_{C}=\frac{\bar{S}_{C}}{\bar{S}_{E}}=1948.25$$

$$F_{A\times B}=\frac{\bar{S}_{A\times B}}{\bar{S}_{E}}=8.25$$

$$F_{A\times C}=\frac{\bar{S}_{A\times C}}{\bar{S}_{E}}=1.75$$

$$F_{B\times C}=\frac{\bar{S}_{B\times C}}{\bar{S}_{E}}=1.25$$

5、显著性检验

查 F 分布表：

$F_{0.01}(2,8)$＝8.65；$F_{0.05}(2,8)$＝4.46；$F_{0.1}(2,8)$＝3.28；$F_{0.01}(4,8)$＝7.01；$F_{0.05}(4,8)$＝3.84；$F_{0.1}(4,8)$＝2.81

因为 F_A、F_B、F_C 均大于 $F_{0.01}(2,8)$＝8.65，$F_{A\times B}$ 大于 $F_{0.01}(4,8)$＝7.01，$F_{A\times C}$ 和 $F_{B\times C}$ 均小于 $F_{0.1}(4,8)$＝2.81，所以因素 *A* 反应温度、因素 *B* 反应压力、因素 *C* 溶液浓度以及因素 *A* 与因素 *B* 的交互作用对试验指标产品产量均有高度显著性影响，因素 *A* 与因素 *C* 的交互作用和因素 *B* 与因素 *C* 的交互作用对试验指标产品产量没有影响。

6、列出方差分析表

方差分析表

离差来源	离差平方和	自由度	均方	统计量	临界值	显著性
A	2.04	2	1.02	25.50		***
B	1.17	2	0.58	14.50	$F_{0.01}(2,8)$＝8.65	***
A×B	1.32	4	0.33	8.25	$F_{0.05}(2,8)$＝4.46	***
C	155.87	2	77.93	1948.25	$F_{0.1}(2,8)$＝3.28	***
A×C	0.28	4	0.07	1.75	$F_{0.01}(4,8)$＝7.01	
B×C	0.18	4	0.05	1.25	$F_{0.05}(4,8)$＝3.84	
误差	0.34	8	0.04		$F_{0.1}(4,8)$＝2.81	
总离差	161.20	26				

4. 重复试验与重复取样正交试验结果的方差分析

1、重复试验与重复取样的基本概念

2、正交试验做重复试验（取样）的原因

① 正交表各列已被因素及交互作用占满，没有空白列也无经验误差（由以往的经验确定），这时为了估计试验误差进行方差分析，一般除选用更大的正交表外，还可作重复试验（取样）。

② 虽然因素没有占满正交表的所有列，即尚有少数空白列，但由于试验的原因做了重复试验（取样）（为了提高试验精度，减少试验误差的干扰）。

3、误差平方和的分类及使用

1）分类

(1) 第一类误差：从正交表空白列计算出来的误差Se_1称为第一类误差。

其自由度 fe_1＝正交表的空白列自由度相加

(2) 第二类误差：将不同的条件的重复试验（取样）所得得试验结果内部的偏差平方和汇总得到Se_2。

其自由度 $fe_2=r(k-1)$

其中k为各号试验的重复次数，r为试验号总数。

2）使用方法:

(1) 重复试验

① 正交表上无空白列时，Se_1不存在，则$Se=Se_2$；$fe=fe_2$

② 正交表上有空白列时，Se_1 存在，则$Se=Se_1+Se_2$；$fe=fe_1+fe_2$

(2) 重复取样

① 正交表上无空白列时Se_1不存在，用Se_2代替Se进行检验，

此时fe＝fe_2，若检验结果有一半以上的因素及交互作用不显著时，用Se_2代替Se是合理的，否则用Se_2代替Se是不合理的(注：这是由于Se_2较小，不能用它来检验各因素水平之间是否存在差异，若要进行方差分析，必需选用大的正交表或者重新做试验)。

② 正交表上有空白列时 S_{e1} 存在，计算 $F_{比}=\dfrac{S_{e1}/f_{e1}}{S_{e2}/f_{e2}}$

当 $F_{比} \geqslant F_{0.05}(f_{e1}, f_{e2})$ 时，舍去 S_{e2}，此时 $S_e = S_{e1}$；$f_e = f_{e1}$

当 $F_{比} < F_{0.05}(f_{e1}, f_{e2})$ 时，$S_e = S_{e1} + S_{e2}$；$f_e = f_{e1} + f_{e2}$

实例：硅钢带取消空气退火工艺试验，空气退火能脱除一部分碳，但同时钢带表面会生成一层很厚的氧化皮，增加酸洗的困难。欲取消这道工序，为此要做试验，试验指标是钢带的磁性，看一看取消这道工序后钢带磁性有没有大的变化。本试验考察的因素及水平下表，试通过正交试验设计及方差分析确定是否可以取消空气退火这道工序。

因素及水平表

因素水平	A 退火工艺	B 成品厚度 mm
1	空气退火	0.20
2	取消空气退火	0.35

正交试验安排及试验结果表

表头设计 列号 试验号	A	B		试验结果 x_{jk}=(原数×1/100-184) j=1,2,3,4;k=1,2,3,4,5					合计
	1	2	3						
1	1	1	1	2.0	5.0	1.5	2.0	1.0	11.5
2	1	2	2	8.0	5.0	3.0	7.0	2.0	25.0
3	2	1	2	4.0	7.0	0	5.0	6.5	22.5
4	2	2	1	7.5	7.0	5.0	4.0	1.5	25.0
K_1	36.5	34.0	36.5						
K_2	47.5	50.0	47.5						
$\overline{K}_1$	18.25	17.0	18.25						
$\overline{K}_2$	23.75	25.0	23.75						
R	5.5	8.0	5.5						
S_i	6.05	12.8	6.05						

试验结果的方差分析

1、计算离差平方和

1）计算总离差平方和

$$S_T=\sum_{i=1}^{r}\sum_{j=1}^{k}(x_{ij}-\bar{x})^2=\sum_{i=1}^{4}\sum_{j=1}^{5}x_{ij}^2-20\bar{x}^2=115.2$$

其中

$$\bar{x}=\frac{1}{rk}\sum_{i=1}^{r}\sum_{j=1}^{k}x_{ij}=4.2$$

其中：r 表示不同条件的试验个数；

k 表示每个试验条件下的重复试验次数；

x_{ij} 表示第 i 次试验条件下的第 j 次重复试验结果。

这里 S_T应由三部分组成：因素离差平方和（条件变差）；第一类误差；第二类误差。

2）计算因素的离差平方和

$$S_i=\frac{1}{ak}\sum_{j=1}^{p}x_{ij}^{\ 2}-rk\bar{x}^2=\frac{1}{10}\sum_{j=1}^{2}x_{ij}^2-20\bar{x}^2$$

$$S_1=\frac{1}{10}\sum_{j=1}^{2}x_{1j}^2-20\bar{x}^2=(36.5^2+47.5^2)/10-20\times4.2^2=6.05$$

$$S_2=\frac{1}{10}\sum_{j=1}^{2}x_{2j}^2-20\bar{x}^2=(34^2+50^2)/10-20\times4.2^2=12.80$$

$$S_3=\frac{1}{10}\sum_{j=1}^{2}x_{3j}^2-20\bar{x}^2=(36.5^2+47.5^2)/10-20\times4.2^2=6.05$$

由此得出：

$S_A=S_1=6.05$；$S_B=S_2=12.8$；$S_{E1}=S_3=6.05$

3）计算第二类误差及误差平方和

我们知道在此例中存在两种误差，第一和第二类误差，第一类误差已经求出，下面我们来求第二类误差。

第二类误差有两种求法

（1）根据第二类误差的定义，同一号试验我们重复做了 5 次，5 个数据间的差异反映了随机误差的影响。因此如果用 $\bar{x}_i$ 表示第 i 号试验的 5 个数据的算术平均值（$\bar{x}_i=\frac{1}{5}\sum_{j=1}^{5}x_{ij}$）那么第 i 号试验的 5 个数据的内部误差平方和。

$\Delta_i=\sum_{j=1}^{5}(x_{ij}-\bar{x}_i)$；所以 $S_{E2}=\sum_{i=1}^{4}\Delta_i=90.3$

(2)令 $S_T'=k\sum_{i=1}^{4}(\bar{x}_i-\bar{x})^2=24.9$

它相当于以前我们讲过的没有重复试验的正交试验设计中的总离差平方和，联系我们讲过的内容，它应该包括两部分，一部分是第一类误差，另一部分是因素离差平方和（条件变差），由于我们已经求出了总离差平方和，它包括三部分：第一类误差；第二类误差和因素离差平方和。

所以 $S_{E2} = S_T - S_T' = 90.3$

$S_E = S_{E1} + S_{E2} = 96.35$

2、计算自由度

1）f_T=总的试验次数－1＝20-1＝19

2）f_A=f_B =水平数－1＝2-1＝1

3）f_{E1}= f_3＝2-1＝1

4）f_{E2}＝$r(k-1)$ ＝4×4＝16

5）f_E＝f_{E1}＋f_{E2}＝16＋1＝17

3、计算平均离差平方和

$$\overline{S}_A = \frac{S_A}{f_A} = 6.05;\quad \overline{S}_B = \frac{S_B}{f_B} = 12.8;\quad \overline{S}_E = \frac{S_E}{f_A} = 5.67$$

4、求 $F_{比}$

$$F_A = \frac{\overline{S}_A}{\overline{S}_E} = 1.067;\quad F_B = \frac{\overline{S}_B}{\overline{S}_E} = 2.257$$

5、显著性检验

查 F 分布表：

$F_{0.01}(1,17) = 8.40$； $F_{0.05}(1,17) = 4.45$； $F_{0.1}(1,17) = 3.03$

F_A, F_B 均小于 $F_{0.1}(1,17)=3.03$，所以说 A 和 B 均不显著。

6、列出方差分析表

方差分析表

离差来源	离差平方和	自由度	均方	统计量	临界值	显著性
A	6.05	1	6.05	1.067	$F_{0.01}(1,17)=8.40$ $F_{0.05}(1,17)=4.45$ $F_{0.1}(1,17)=3.03$	
B	12.8	1	12.8	2.257		
E_1	6.05	1				
E_2	90.3	16				
E	96.35	17	5.67			
总离差	115.2	19				

5. 正交试验结果的综合分析

实例：煤炭灰化条件的试验，需要考察升温方式（A）、通风方式（B），最终温度（C）对灰化结果的影响，每个因素考察二个水平。通过试验找出各因素最佳水平搭配，并确定各因素对灰化结果影响的主次顺序。

具体试验方案设计如下：

(一) 根据设计任务明确本次试验目的是为了确定最佳灰化条件，确定用灰分做灰化优劣的指标。如灰分越低灰化越完善，其效果越好。

(二) 参考有关资料，确定

因素及水平表

因素水平	A 升温方式	B 通风方式	C 最终温度
1	二次升温法	炉门密闭	750^{+10}_{-20}℃
2	一次升温法	炉门打开 2mm	800^{+20}_{-20}℃

（三）经研究分析，除考察上述三个因素外，还要考察交互作用 $A\times B$、$B\times C$、$A\times C$、$A\times B\times C$，因素及其交互作用共需占据正交表 7 列。选取 2 水平的正交表 $L_8(2^7)$ 最合适，并按其二列间交互作用列表安排试验，进行表头设计，表头设计结果见下表。

试验用正交表及试验结果

表头 \ 试验号	A	B	$A\times B$	C	$A\times C$	$B\times C$	$A\times B\times C$	试验结果 x_{ij}		
列号	1	2	3	4	5	6	7	重复试验 1	重复试验 2	合计
1	1	1	1	1	1	1	1	20.44	20.20	40.64
2	1	1	1	2	2	2	2	20.42	20.32	40.74
3	1	2	2	2	2	1	1	20.18	20.35	40.53
4	1	2	2	2	2	1	1	19.93	20.03	39.96
5	2	1	2	1	2	1	2	20.72	20.55	41.27
6	2	1	2	2	1	2	1	20.58	20.42	41.00
7	2	2	1	1	2	2	1	20.50	20.43	40.93
8	2	2	1	2	1	1	2	20.18	20.12	40.30
K_1	161.87	163.65	162.61	163.37	162.47	162.17	162.53			
K_2	163.50	161.72	162.76	162.00	162.90	163.20	162.84			
$\overline{K}_1$	20.23	20.46	20.33	20.42	20.31	20.27	20.32			
$\overline{K}_2$	20.44	20.22	20.35	20.25	20.36	20.40	20.36			
R	1.63	-1.93	0.15	-1.37	0.43	1.03	0.31			
S_i	0.1660	0.2328	0.0014	0.1173	0.0116	0.0663	0.0060			

(四) 根据表头设计，把各因素的水平取值按因素所占下交表中列号下的水平填入，就构成了正交试验设计的方案。应当注意，试验方案确定后，必须严格按照各号试验条件进行试验，确保正交试验结果的“综合可比性”对试验所提出的要求。

试验操作必须按操作规范执行，以保证试验数据的精度。试验顺序可依次进行，但最好利用随机化数表或抽签使各号试验随机进行。

(五) 正交试验结果的计算分析

试验结果的计算和分析工作对正交试验设计上极为重要的，它可以解决如下几个问题：

1、确定因素对试验指标影响的主次顺序；

2、找出试验条件优化组合方案；

3、分析因素与指标的关系，找出指标随因素变化的规律和趋势，用以指导进一步改进试验方向。

根据试验方案，按各号的试验条件实施试验，得到的试验结果及正交表计算分析试结果填入正交表。

(1) 直观分析

① 试验结果的直接对比

通过对正交表中各试验条件的试验指标的合计的直接对比，$A_1B_2C_2$灰化条件为最好。

② 因素水平效果的统计计算和综合对比（略讲）

③因素效应的直观对比

极差的大小，能反映试验中各因素作用的大小。根据正交表中极差的大小可以得出三个因素的主次顺序（从大到小）为：BAC

（2）方差分析

① 计算离差平方和

$$S_T=\sum_{i=1}^{r}\sum_{j=1}^{k}(x_{ij}-\bar{x})^2=\sum_{i=1}^{8}\sum_{j=1}^{2}x_{ij}^2-16\bar{x}^2=0.686194$$

其中

$$\bar{x}=\frac{1}{rk}\sum_{i=1}^{r}\sum_{j=1}^{k}x_{ij}=20.335625$$

其中：r 表示不同条件的试验个数；

k 表示每个试验条件下的重复试验次数；

x_{ij} 表示第 i 次试验条件下的第 j 次重复试验结果。

$$S_i=\frac{1}{ak}\sum_{j=1}^{p}x_{ij}^{2}-rk\bar{x}^{2}=\frac{1}{8}\sum_{j=1}^{2}x_{ij}^{2}-16\bar{x}^{2}$$

$$S_1=S_A=\frac{1}{8}\sum_{j=1}^{2}x_{1j}^{2}-16\bar{x}^{2}=(161.87^2+163.5^2)/8-16\times 20.335625^2=0.166056$$

同理可以计算得出：

$$S_B=S_2=0.232806$$

$$S_{A\times B}=S_3=0.0014065$$

$$S_C=S_4=0.1173065$$

$$S_{A\times C}=S_5=0.011557$$

$$S_{B\times C}=S_6=0.066307$$

$$S_{A\times B\times C}=S_7=0.0060065$$

② 计算试验误差平方和

直接用重复试验结果，求取试验随机误差 S_{E2}。

$$S_{E2}=S_T-S_{因+交}=0.686194-0.60144=0.08475$$

③ 计算自由度

$$f_T=n-1=8\times 2-1=15$$

$$f_A=f_B=f_C=f_{A\times B}=f_{A\times C}=f_{B\times C}=f_{A\times B\times C}=1$$

$$f_{E2}=f_T-f_A-f_B-f_C-f_{A\times B}-f_{A\times C}-f_{B\times C}-f_{A\times B\times C}=8$$

④ 计算均方

$$\bar{S}_A=\frac{S_A}{f_A}=0.166056$$

同理可以计算出

$\bar{S}_B=0.232806$；$\bar{S}_C=0.1173065$；$\bar{S}_{A\times B}=0.0014065$；$\bar{S}_{B\times C}=0.0014065$；

$\bar{S}_{A\times B\times C}=0.0060065$；$\bar{S}_{E2}=0.01059$

⑤ 计算统计量

$$F_A=\frac{\bar{S}_A}{\bar{S}_{E2}}=15.68$$

同理可以计算出：

F_B=21.98；$F_{A\times B}$=0.13；F_C=11.08；$F_{A\times C}$=1.09；$F_{B\times C}$=6.26；$F_{A\times B\times C}$=0.58

⑥ 显著性检验

初步检验各因素的显著性：

查 F 分布表得 $F_{0.01}(1,8)$=11.3，$F_{0.05}(1,8)$=5.32，$F_{0.1}(1,8)$=3.46。经检验结果，$F_{A\times B}$，$F_{A\times C}$，$F_{A\times B\times C}$ 均小于 $F_{0.1}(1,8)$，所以 A、B、C 三因素的交互作用 $A\times B$、$A\times C$、$A\times B\times C$ 均不显著。为了提高各项效应检验的精度，可将这几项合并成由正交试验分离出的

试验列的离差平方和 S_{E1}，作为试验误差。

$$S_{E1}=S_3+S_5+S_7=0.0014065+0.11557+0.0060056=0.01897$$

$$f_{E1}=f_3+f_5+f_7=3$$

归并两类试验误差，求出作为本次试验检验显著性的试验误差平方和 S_E 及自由度 f_E。

$$S_E=S_{E1}+S_{E2}=0.01897+0.08475=0.10372$$

$$f_E=f_{E1}+f_{E2}=3+8=11$$

⑦列出方差分析表 3-16

离差来源		平方和 S_i	自由度 f	均方 $\bar{S}_i$	统计量 F			
					$\frac{\bar{S}_i}{\bar{S}_{E2}}$	$\frac{\bar{S}_i}{\bar{S}_E}$	$\frac{\bar{S}_i}{\bar{S}_{E1}}$	$\frac{\bar{S}_i}{\bar{S}_{(E1+B\times C)}}$
(1)		(2)	(3)	(4)	(5)	(6)	(7)	(8)
试验因素	A	0.166056	1	0.166056	15.68	17.16	26.26	7.79
	B	0.232806	1	0.232806	21.98	24.69	36.82	10.92
	C	0.1173065	1	0.1173065	11.08	12.44	19.55	5.50
交互作用	B×C	0.066307	1	0.066307	6.26	7.03	10.49	
误差因素	重复试验 E2	0.08475	8	0.01059				
	并入因子 E1	0.01897	3	0.0063233				
	小计	0.10372	11	0.009429				
试验条件组合因素		0.60144	7	0.08592				
总离差		0.686194	15					

$F_{0.01}(1,8)=11.3$，$F_{0.05}(1,8)=5.32$，$F_{0.1}(1,8)=3.46$，$F_{0.01}(1,11)=9.65$，$F_{0.05}(1,11)=4.84$，$F_{0.1}(1,11)=3.23$，$F_{0.01}(1,3)=21.2$，$F_{0.05}(1,3)=10.1$，$F_{0.1}(1,3)=5.54$，$F_{0.01}(1,4)=34.2$，$F_{0.05}(1,4)=7.71$，$F_{0.1}(1,4)=5.46$

表中给出了四种估计试验误差的方法，得出了四组 F 统计量，如表中(5)、(6)、(7)、(8)栏的数据。对比以后可以看出：(5)、(6)、(7)栏的统计检验结论是一致的。也就说，在安排有重复试验的正交试验中，可用交互作用不显著或可忽略的因子合并来估计试验误差，据此统计检验得出的结论还是可靠的，但若把不可忽略交互作用的因子并入试验误差如表 3-45 第(8)栏，其统计检验的结论与实际情况(第 5、6 栏得出的结论)差距较大，甚至造成错误的判断。这是采用方差分析必须注意的问题。

6. 正交试验设计结果的效应计算与指标值的预估计

1、正交试验设计试验数据结构

在正交试验设计中，若以 μ_t 表示第 t 号试验各因素水平搭配对指标值 x_t 影响的总和，也叫 x_t 的理论值，以 ε_t 表示第 t 号试验的随机误差，则有

$$x_t = \mu_t + \varepsilon_t \qquad (t=1,2,3,\ldots n)$$

$\varepsilon_t \sim N(0,\sigma^2)$，相互独立

这种数据结构式称为在 $L_n(m^k)$ 型正交表上安排试验的数学模型。由于各正交表的具体情况不同，数据结构的具体形式也不同。下面对几个常用的正交表，分别写出它们的数据结构式。

1） $L_4(2^3)$ 表上的数据结构

假设安排两个因素 A、B，A 安排在第 1 列，B 安排在第 2 列，根据各因素水平组合的不

同，数据结构的形式如下表所示。

$L_4(2^3)$ 正交表的数据结构形式

试验号 \ 因素 / 列号	A	B		数据结构
	1	2	3	
1	1	1	1	$x_1=\mu_{11}+\varepsilon_1$
2	1	2	2	$x_2=\mu_{12}+\varepsilon_2$
3	2	1	2	$x_3=\mu_{21}+\varepsilon_3$
4	2	2	1	$x_4=\mu_{22}+\varepsilon_4$

表中，μ_{ij} 表示在 A_i，B_j 组合下指标值 x_i 的理论值。在不同情况下，μ_{ij} 的具体形式也不同。

（1）不考虑交互作用

$$\mu_{ij}=\mu+a_i+b_j \qquad (i=1,2;\ j=1,2)$$

其中，$\mu=\frac{1}{4}(\mu_{11}+\mu_{12}+\mu_{21}+\mu_{22})$，称为总平均（工程平均值）

a_i为因素 A 在 i 水平时的效应，$a_1+a_2=0$，

b_j为因素 B 在 j 水平时的效应，$b_1+b_2=0$，

数据结构形式如下表所示。

无交互作用时 $L_4(2^3)$ 正交表的数据结构形式

因素 列号 试验号	A	B		数据结构
	1	2	3	
1	1	1	1	$x_1=\mu_{11}+\varepsilon_1=\mu+a_1+b_1+\varepsilon_1$
2	1	2	2	$x_2=\mu_{12}+\varepsilon_2=\mu+a_1+b_2+\varepsilon_2$
3	2	1	2	$x_3=\mu_{21}+\varepsilon_3=\mu+a_2+b_1+\varepsilon_3$
4	2	2	1	$x_4=\mu_{22}+\varepsilon_4=\mu+a_2+b_2+\varepsilon_4$

（2）考虑交互作用

$$\mu_{ij}=\mu+a_i+b_j+(ab)_{ij} \qquad （i=1,2；j=1,2）$$

其中，$(ab)_{ij}$为 A_i，B_j组合下交互作用的效应。

$(ab)_{11}+(ab)_{21}=0$，$(ab)_{12}+(ab)_{22}=0$

数据结构形式如下表所示

有交互作用时 $L_4(2^3)$ 正交表的数据结构形式

因素 列号 试验号	A 1	B 2	 3	数据结构
1	1	1	1	$x_1=\mu_{11}+\varepsilon_1=\mu+a_1+b_1+(ab)_{11}+\varepsilon_1$
2	1	2	2	$x_2=\mu_{12}+\varepsilon_2=\mu+a_1+b_2+(ab)_{12}+\varepsilon_2$
3	2	1	2	$x_3=\mu_{21}+\varepsilon_3=\mu+a_2+b_1+(ab)_{21}+\varepsilon_3$
4	2	2	1	$x_4=\mu_{22}+\varepsilon_4=\mu+a_2+b_2+(ab)_{22}+\varepsilon_4$

2） **$L_9(3^4)$ 表上的数据结构**

（1）不考虑交互作用

假设安排两个因素 A、B，A 安排在第 1 列，B 安排在第 2 列，根据各因素水平组合的不同，数据结构的形式如下表所示。

无交互作用时 $L_9(3^4)$ 正交表的数据结构形式

表头设计	A	B			数据结构形式
试验号 \ 列号	1	2	3	4	
1	1	1	1	1	$x_1=\mu_{11}+\varepsilon_1=\mu+a_1+b_1+\varepsilon_1$
2	1	2	2	2	$x_2=\mu_{12}+\varepsilon_2=\mu+a_1+b_2+\varepsilon_2$
3	1	3	3	3	$x_3=\mu_{13}+\varepsilon_3=\mu+a_1+b_3+\varepsilon_3$
4	2	1	2	3	$x_4=\mu_{21}+\varepsilon_4=\mu+a_2+b_1+\varepsilon_4$
5	2	2	3	1	$x_5=\mu_{22}+\varepsilon_5=\mu+a_2+b_2+\varepsilon_5$
6	2	3	1	2	$x_6=\mu_{23}+\varepsilon_6=\mu+a_2+b_3+\varepsilon_6$
7	3	1	3	2	$x_7=\mu_{31}+\varepsilon_7=\mu+a_3+b_1+\varepsilon_7$
8	3	2	1	3	$x_8=\mu_{32}+\varepsilon_8=\mu+a_3+b_2+\varepsilon_8$
9	3	3	2	1	$x_9=\mu_{33}+\varepsilon_9=\mu+a_3+b_3+\varepsilon_9$

上表中的数据结构式中：

$$x_{ij}=\mu+a_i+b_j+\varepsilon_{ij}$$

$$\varepsilon_{ij}\sim N(0,\sigma^2)$$ ，相互独立

$$\sum_{i=1}^{3}a_i=0\text{，}\qquad \sum_{j=1}^{3}b_j=0$$

（2）考虑交互作用

假设安排两个因素 A、B，A 安排在第 1 列，B 安排在第 2 列，则 A 与 B 交互作用应安排在 3、4 两列。根据各因素水平组合的不同，数据结构的形式如下表所示。

考虑交互作用时 $L_9(3^4)$ 正交表的数据结构形式

表头设计 / 列号 / 试验号	A	B	A×B	A×B	数据结构形式
	1	2	3	4	
1	1	1	1	1	$x_1=\mu_{11}+\varepsilon_1=\mu+a_1+b_1+(ab)_{11}^1+(ab)_{11}^2+\varepsilon_1$
2	1	2	2	2	$x_2=\mu_{12}+\varepsilon_2=\mu+a_1+b_2+(ab)_{12}^1+(ab)_{12}^2+\varepsilon_2$
3	1	3	3	3	$x_3=\mu_{13}+\varepsilon_3=\mu+a_1+b_3+(ab)_{13}^1+(ab)_{13}^2+\varepsilon_3$
4	2	1	2	3	$x_4=\mu_{21}+\varepsilon_4=\mu+a_2+b_1+(ab)_{21}^1+(ab)_{21}^2+\varepsilon_4$
5	2	2	3	1	$x_5=\mu_{22}+\varepsilon_5=\mu+a_2+b_2+(ab)_{22}^1+(ab)_{22}^2+\varepsilon_5$
6	2	3	1	2	$x_6=\mu_{23}+\varepsilon_6=\mu+a_2+b_3+(ab)_{23}^1+(ab)_{23}^2+\varepsilon_6$
7	3	1	3	2	$x_7=\mu_{31}+\varepsilon_7=\mu+a_3+b_1+(ab)_{31}^1+(ab)_{31}^2+\varepsilon_7$
8	3	2	1	3	$x_8=\mu_{32}+\varepsilon_8=\mu+a_3+b_2+(ab)_{32}^1+(ab)_{32}^2+\varepsilon_8$
9	3	3	2	1	$x_9=\mu_{33}+\varepsilon_9=\mu+a_3+b_3+(ab)_{33}^1+(ab)_{33}^2+\varepsilon_9$

在表 4-62 的数据结构中：

$$x_{ij}=\mu+a_i+b_j+(ab)_{ij}^1+(ab)_{ij}^2+\varepsilon_{ij}$$

$$\sum_{i=1}^{3}a_i=0,\quad \sum_{j=1}^{3}b_j=0,\quad \sum_{i=1}^{3}(ab)_{ij}=0,\quad \sum_{j=1}^{3}(ab)_{ij}=0,\quad \varepsilon_{ij}\sim N(0,\sigma^2)$$，相互独立

（$i,j=1,2,3$）

2、正交试验设计中的效应计算

由下式：

$$x_t = \mu_t + \varepsilon_t \qquad (t = 1,2,3,\cdots,n)$$

ε_t 是随机误差，我们要对 μ_t 作出估计，即求出 μ_t 的估计值$\hat{\mu}_t$，使得满足：

$$\sum_{i=1}^{n}(x_t - \hat{\mu}_t)^2 = \min\sum_{i=1}^{n}(x_t - \mu_t)^2$$

下面以 $L_4(2^3)$ 表安排两个因素 A，B，且以考虑交互作用的情况为例说明这个问题，由 $L_4(2^3)$ 表上的数据结构式可得：

残差平方和为

$$\begin{aligned}S &= \sum_{i=1}^{n}(x_t - \mu_t)^2 \\ &= [x_1 - \mu - a_1 - b_1 - (ab)_{11}]^2 \\ &\quad + [x_2 - \mu - a_1 - b_2 - (ab)_{12}]^2 \\ &\quad + [x_3 - \mu - a_2 - b_1 - (ab)_{21}]^2 \\ &\quad + [x_4 - \mu - a_2 - b_2 - (ab)_{22}]^2\end{aligned}$$

为了使 S 达到最小，采用最小二乘法，求 μ_t 的估计值$\hat{\mu}_t$，就是求 $a_i, b_j, (ab)_{ij}$的估计值$\hat{\mu}, \hat{a}_i, \hat{b}_j, \widehat{ab}_{ij}$.

令

$$\begin{aligned}\frac{\partial S}{\partial \mu} = &-2[x_1 - \mu - a_1 - b_1 - (ab)_{11}] \\ &-2[x_2 - \mu - a_1 - b_2 - (ab)_{12}] \\ &-2[x_3 - \mu - a_2 - b_1 - (ab)_{21}]\end{aligned}$$

$$-2[x_4-\mu-a_2-b_2-(ab)_{22}]$$
$$=0$$

由于 $a_1+a_2=0, b_1+b_2=0, (ab)_{11}+(ab)_{21}=0, (ab)_{12}+(ab)_{22}=0$,所以有:

$$x_1+x_2+x_3+x_4-4\mu=0$$

得出:

$$\mu=\frac{1}{4}(x_1+x_2+x_3+x_4)$$

记为:

$$\hat{\mu}=\bar{x}$$

令

$$\frac{\partial S}{\partial a_1}=-2[x_1-\mu-a_1-b_1-(ab)_{11}]$$
$$-2[x_2-\mu-a_1-b_2-(ab)_{12}]$$
$$=0$$

由于 $b_1+b_2=0, (ab)_{11}+(ab)_{12}=0$,所以有:

$$x_1+x_2-2\mu-2a_1=0$$

得出:

$$a_1=\frac{x_1+x_2}{2}-\mu=\bar{A}_1-\mu$$

上式中,$\bar{A}_1$ 为 A 因素 1 水平指标值的平均值. 记为:

$$\hat{a}_1=\bar{A}_1-\hat{\mu}$$

同理可得出:

$$\hat{a}_2=\bar{A}_2-\hat{\mu}$$

$$\hat{b}_1 = \overline{B}_1 - \hat{\mu}$$

$$\hat{b}_2 = \overline{B}_2 - \hat{\mu}$$

上式中,$\overline{A}_2$,$\overline{B}_1$,$\overline{B}_2$ 为各因素水平所对应的指标平均值.

令
$$\frac{\partial S}{\partial (ab)_{11}} = -2[x_1 - \mu - a_1 - b_1 - (ab)_{11}] = 0$$

得出:

$$(ab)_{11} = x_1 - a_1 - b_1 - \mu$$

记为:

$$(\hat{ab})_{11} = \overline{A_1 B_1} - \hat{a}_1 - \hat{b}_1 - \hat{\mu}$$

上式中,$\overline{A_1 B_1}$表示 $A_1 B_1$ 水平搭配的指标平均值.

同理可得出:

$$(\hat{ab})_{12} = \overline{A_1 B_2} - \hat{a}_1 - \hat{b}_2 - \hat{\mu}$$

$$(\hat{ab})_{21} = \overline{A_2 B_1} - \hat{a}_2 - \hat{b}_1 - \hat{\mu}$$

$$(\hat{ab})_{22} = \overline{A_2 B_2} - \hat{a}_2 - \hat{b}_2 - \hat{\mu}$$

上式中,$\overline{A_1 B_2}$,$\overline{A_2 B_1}$,$\overline{A_2 B_2}$表示各因素水平搭配的指标平均值.

对其他的正交表,针对它们的数学模型,采用同样的方法,就能得出相应的效应值,从

一般的效应计算公式：

$$\hat{a}_i = \overline{A}_i - \overline{x}$$

$$\hat{b}_j = \overline{B}_j - \overline{x}$$

$$\hat{c}_k = \overline{C}_k - \overline{x}$$

$$\hat{d}_l = \overline{D}_l - \overline{x}$$

………

$$(\widehat{ab})_{ij} = \overline{A_iB_j} - \hat{a}_i - \hat{b}_j - \overline{x}$$

$$(\widehat{ac})_{ik} = \overline{A_iC_k} - \hat{a}_i - \hat{c}_k - \overline{x}$$

$$(\widehat{bc})_{jk} = \overline{B_jC_k} - \hat{b}_j - \hat{c}_k - \overline{x}$$

………

7. 小结

1、方差分析是正交试验设计结果行之有效的统计分析手段；

2、掌握方差分析基本原理是理解方差分析的基础；

3、直观分析快速、方便，但也存在可靠性相对较差和不能适用于所有类型正交试验的缺点。

4、对一些不显著的因素，可以将其方差并入误差，以提高检验的精度；

5、对空白列的方差分析同样要给予重视，它是判断交互作用是否存在及是否漏掉重要因素的一个窗口；

6、对有重复试验（取样）的正交试验设计结果的方差分析要特别注意第二类误差及其使用；

7、对取得的最佳组合在条件允许的情况尽可能进行验证试验；

8、根据因素水平变化趋势确定是否需要重新调整因素水平补加试验

操千曲而后晓声

观千剑而后识器

你的进步，我的快乐！