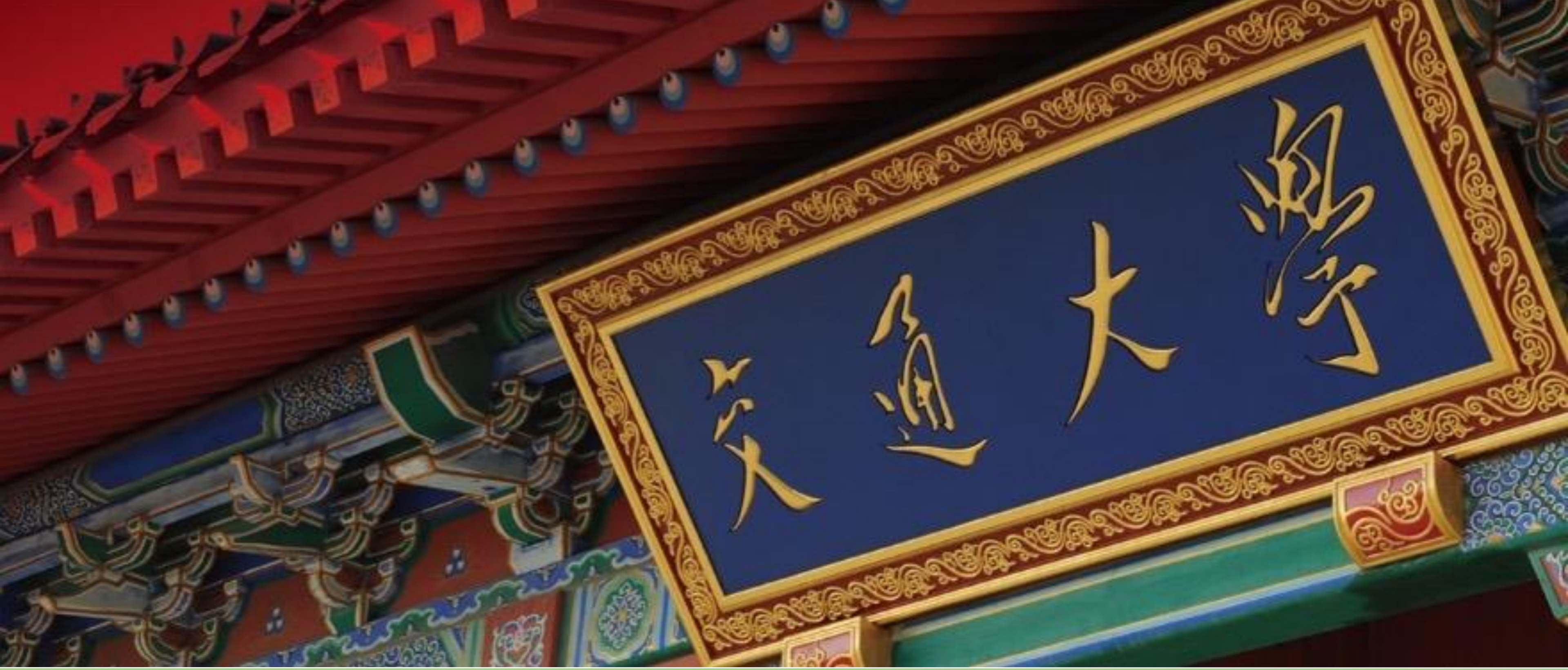

基础数理统计（研究生公共课）

肖柳青 博士 教授 主讲

上海交通大學
SHANGHAI JIAO TONG UNIVERSITY

第 五章 方差分析

（ANOVA--The Analysis of Variance）

主要内容

5. 1　单因素方差分析

5. 2　双因素方差分析

5. 3　正交试验设计

§ 5. 2* 双因素方差分析

(Two-Way ANOVA)

5.2.1 无交互作用

5.2.2 有交互作用

双因素方差分析是讨论两因素试验的统计推断问题。分无交互作用（非重复试验）和有交互作用（重复试验）两种情形进行讨论。

5.2.1 无交互作用

一、两因素非重复无交互作用的方差分析

数学模型

设有两个因素 A，B，因素 A 有 r 个不同水平：A_1，$A_2,\cdots$，A_r；因素 B 有 s 个不同水平：B_1，$B_2,\cdots$，B_s，在 A，B 的每一种组合水平 (A_i,B_j) 下作一次试验，试验结果为 $X_{ij}=(i=1,\cdots,r,j=1,2,\cdots,s)$，所有 X_{ij} 相互独立，这样共得 rs 个试验结果（表 5.8）。

表5. 8

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

这种对每个组合水平$(A_i, B_j)(i=1,2,\cdots,r, j=1,2,\cdots,s)$各作一次试验的情形称为两因素非重复试验。

例:　用3种电烤箱烧烤3种菜肴，
考察用电量（千瓦小时）

Menu Day 菜肴	Range 1 烤箱 1	Range 2 烤箱 2	Range 3 烤箱 3
1	3.97	4.24	4.44
2	2.39	2.61	2.82
3	2.76	2.75	3.01

假定总体X_{ij}服从正态分布$N(\mu_{ij},\sigma^2)$，其中

$$\mu_{ij}=\mu+\alpha_i+\beta_j,\quad i=1,2,\cdots,r,j=1,2,\cdots,s, \qquad (5.2.4)$$

而$\sum_{i=1}^{r}a_i=0,\qquad \sum_{j=1}^{s}\beta_j=0$。 （5. 2. 3）

于是X_{ij}可表示为

$$\begin{cases}X_{ij}=\mu+\alpha_i+\beta_j+\varepsilon_{ij},\\ \varepsilon_{ij}\sim N(0,\sigma^2),\end{cases} i=1,2,\cdots,r;j=1,2,\cdots,s, \qquad (5.2.7)$$

其中诸ε_{ij}相互独立，α_i称为因素 A在水平A_i引起的效应，它表示水平A_i在总体平均数上引起的偏差。

同样，β_j称为因素B在水平B_j引起的效应，它表示水平B_j在总体平均数上引起的偏差。

所以要推断因素A的影响是否显著，就等价于要检验假设

$$H_{01}: \ \alpha_1 = \alpha_2 = \cdots \alpha_r = 0 \leftrightarrow H_{11};$$

至少有一个

$$\alpha_i \neq 0, i = 1, \cdots, r。$$

类似地，要推断因素B的影响是否显著，就等价于要检验假设

$$H_{02}: \quad \beta_1 = \beta_2 = \cdots = \beta_r = 0 \leftrightarrow H_{12};$$

至少有一个

$$\beta_j \neq 0, j = 1, \cdots, s。$$

当H_{01}成立时，从式（5.2.4）可以看出，均值 μ_{ij}与 α_i无关，这表明因素A对试验结果无显著影响。

同理，当H_{02}成立时，从式（5.2.4）可以看出，均值μ_{ij}与β_j无关，这表明因素B对试验结果无显著影响，当H_{01}，H_{02}都成立时，$\mu_{ij}=\mu$，X_{ij}的波动主要是由随机因素引起的。

导出检验假设H_{01}与H_{02}统计量的方式与单因素方差分析相类似，可采用离差平方和分解的方法。

记

$$\overline{X}_{i\cdot}=\frac{1}{s}\sum_{j=1}^{s}X_{ij}\qquad (i=1,2,\cdots,r),$$

$$\overline{X}_{\cdot j} = \frac{1}{r}\sum_{i=1}^{r} X_{ij} \qquad (j = 1,2,\cdots,s)$$

$$\overline{X} = \frac{1}{rs}\sum_{i=1}^{r}\sum_{j=1}^{s} X_{ij} = \frac{1}{r}\sum_{i=1}^{r}\overline{X}_{i\cdot} = \frac{1}{s}\sum_{j=1}^{s}\overline{X}_{\cdot j}$$

于是总离差平方和

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

$$= \sum_{i=1}^{r}\sum_{j=1}^{s}[(X_{ij} - \overline{X}_{i\cdot} - \overline{X}_{\cdot j} + \overline{X}) + (\overline{X}_{i\cdot} - \overline{X}) + (\overline{X}_{\cdot j} - \overline{X})]^2$$

$$= \sum_{i=1}^{r}\sum_{j=1}^{s}(X_{ij} - \overline{X}_{i\cdot} - \overline{X}_{\cdot j} + \overline{X})^2 + \sum_{i=1}^{r}\sum_{j=1}^{s}(\overline{X}_{i\cdot} - \overline{X})^2$$

$$+ \sum_{i=1}^{r}\sum_{j=1}^{s}(\overline{X}_{\cdot j} - \overline{X})^2$$

$$+2\sum_{i=1}^{r}\sum_{j=1}^{s}(X_{ij}-\overline{X}_{i}\cdot-\overline{X}\cdot_{j}+\overline{X})(\overline{X}\cdot_{j}-\overline{X})$$

$$+2\sum_{i=1}^{r}\sum_{j=1}^{s}(X_{ij}-\overline{X}_{i}\cdot-\overline{X}\cdot_{j}+\overline{X})(\overline{X}_{i}\cdot-\overline{X})$$

$$+2\sum_{i=1}^{r}\sum_{j=1}^{s}(\overline{X}_{i}\cdot-\overline{X})(\overline{X}\cdot_{j}-\overline{X})$$

$$=s\sum_{i=1}^{s}(\overline{X}_{i}\cdot-\overline{X})^{2}$$

$$+r\sum_{j=1}^{s}(\overline{X}\cdot_{j}-\overline{X})^{2}+\sum_{i=1}^{s}\sum_{j=1}^{s}(X_{ij}-\overline{X}_{i}\cdot-\overline{X}\cdot_{j}+\overline{X})^{2}$$

令

$$\left.\begin{aligned} Q_A &= s\sum_{i=1}^{r}(\overline{X}_{i}.-\overline{X})^2, \\ Q_B &= r\sum_{j=1}^{s}(\overline{X}._{j}-\overline{X})^2, \\ Q_E &= \sum_{i=1}^{r}\sum_{j=1}^{s}(X_{ij}-\overline{X}_{i}.-\overline{X}._{j}+\overline{X})^2, \end{aligned}\right\} \quad (5.2.11)$$

则可得

$$Q = Q_A + Q_B + Q_E$$

上式称为**总离差平方和分解式**。其中Q_A为因素A引起的离差平方和，Q_B为因素B引起的离差平方和，Q_E称为**随机误差平方和**。

为了更清楚地看出各离差平方和的意义，与 $\overline{X}_{i\cdot}, \overline{X}_{\cdot j}, \overline{X}$ 的表达式相类似，引进 $\bar{\varepsilon}_{i\cdot}, \bar{\varepsilon}_{\cdot j}$ 与 $\bar{\varepsilon}$，

$$\bar{\varepsilon}_{i\cdot} = \frac{1}{s}\sum_{j=1}^{s}\varepsilon_{ij}, \qquad i = 1,2,\cdots,r,$$

$$\bar{\varepsilon}_{\cdot j} = \frac{1}{r}\sum_{i=1}^{r}\varepsilon_{ij}, \qquad j = 1,2,\cdots,s,$$

$$\bar{\varepsilon} = \frac{1}{rs}\sum_{i=1}^{r}\sum_{j=1}^{s}\varepsilon_{ij} = \frac{1}{r}\sum_{i=1}^{r}\bar{\varepsilon}_{i\cdot} = \frac{1}{s}\sum_{j=1}^{s}\bar{\varepsilon}_{\cdot j}$$

应用式（5.2.7）可把式（5.2.11）写成

$$
\left.\begin{aligned}
Q_A &= s\sum_{i=1}^{r}(\alpha_i+\bar{\varepsilon}_{i\cdot}-\bar{\varepsilon})^2,\\
Q_B &= r\sum_{j=1}^{s}(\beta_j+\bar{\varepsilon}_{\cdot j}-\bar{\varepsilon})^2,\\
Q_E &= \sum_{i=1}^{r}\sum_{j=1}^{s}(\varepsilon_{ij}-\bar{\varepsilon}_{i\cdot}-\bar{\varepsilon}_{\cdot j}+\bar{\varepsilon})^2,
\end{aligned}\right\}\qquad (5.2.14)
$$

从上式看出Q_A主要依赖于因素A的效应；Q_B主要依赖于B的效应；Q_E依赖于随机误差ε_{ij}. 由于

$$
\varepsilon_{ij}\sim N(0,\sigma^2),\bar{\varepsilon}_{i\cdot}\sim N\left(0,\frac{\sigma^2}{s}\right),\bar{\varepsilon}_{\cdot j}\sim N\left(0,\frac{\sigma^2}{r}\right),
$$

$$
\bar{\varepsilon}\sim N\left(0,\frac{\sigma^2}{rs}\right)\qquad (i=1,2,3,\cdots,r;\ j=1,2,\cdots,s)
$$

从而

$$E(Q_A) = sE\left[\sum_{i=1}^{r}(\alpha_i + \bar{\varepsilon}_i. - \bar{\varepsilon})^2\right]$$

$$= s\sum_{i=1}^{r} a_i^2 + sE\left[\sum_{i=1}^{r}(\bar{\varepsilon}_i. - \bar{\varepsilon})^2 + 2s\sum_{i=1}^{r}\alpha_i E(\bar{\varepsilon}_i. - \bar{\varepsilon})\right]$$

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

同理可得

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

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

令

$$\overline{Q}_A = \frac{1}{r-1}Q_A,\quad \overline{Q}_B = \frac{1}{s-1}Q_B,\quad Q_E = \frac{1}{(r-1)(s-1)}Q_E$$

则有

$$E(\overline{Q}_A) = \sigma^2 + \frac{s}{r-1}\sum_{i=1}^{r}\alpha_i^2,\quad E(\overline{Q}_B) = \sigma^2 + \frac{r}{s-1}\sum_{j=1}^{s}\beta_j^2,\quad E(\overline{Q}_E) = \sigma^2$$

当H_{01}成立时，$E(\overline{Q}_A) = E(\overline{Q}_E)$，否则，$E(\overline{Q}_A) > E(\overline{Q}_E)$，

当H_{02}成立时，$E(Q_B) = E(\overline{Q}_B)$，否则，$E(\overline{Q}_B) > E(\overline{Q}_E)$，

令

$$F_A = \frac{Q_A/(r-1)}{Q_E/(r-1)(s-1)} \stackrel{\text{def}}{=\!=} \frac{\overline{Q}_A}{\overline{Q}_E}$$

$$F_B = \frac{Q_B/(s-1)}{Q_E/(r-1)(s-1)} \stackrel{\text{def}}{=\!=} \frac{\overline{Q}_B}{\overline{Q}_E}$$

当H_{01}，H_{02}不成立时，F_A，F_B有偏大趋势，因此F_A，F_B可作为检验假设H_{01}，H_{02}的统计量。

下面导出检验F_A与F_B的概率分布。当H_{01}及H_{02}成立时，

$$\alpha_i = \beta_j = 0\,(i = 1,2,\cdots,r, j = 1,2,\cdots,s),$$

因而$X_{ij} = \mu + \varepsilon_{ij}$，各离差平方和可改写为

$$Q_A = s\sum_{i=1}^{r}(\bar{\varepsilon}_{i\cdot} - \bar{\varepsilon})^2$$

$$Q_B = r\sum_{j=1}^{s}(\bar{\varepsilon}_{\cdot j} - \bar{\varepsilon})^2$$

$$Q_E = \sum_{i=1}^{r}\sum_{j=1}^{s}(\varepsilon_{ij} - \bar{\varepsilon}_{i\cdot} - \bar{\varepsilon}_{\cdot j} + \bar{\varepsilon})^2$$

$$Q = \sum_{i=1}^{r}\sum_{j=1}^{s}(\varepsilon_{ij} - \bar{\varepsilon})^2 = Q_A + Q_B + Q_E$$

由于$Q = \sum_{i=1}^{r}\sum_{j=1}^{s}\varepsilon_{ij}^2 - rs\bar{\varepsilon}^2$，于是有

$$\sum_{i=1}^{r}\sum_{j=1}^{s}\varepsilon_{ij}^2 = \sum_{i=1}^{r}\sum_{j=1}^{s}(\varepsilon_{ij} - \bar{\varepsilon})^2 + rs\bar{\varepsilon}^2 = Q_A + Q_B + Q_E + rs\bar{\varepsilon}^2$$

为了利用柯赫伦因子分解定理，上式两边同除以σ^2。

于是，等式左边$\frac{1}{\sigma^2}\sum_{i=1}^{r}\sum_{j=1}^{s}\varepsilon_{ij}^2$是自由度为$rs$的$\chi^2$的变量，

而等式右边四项及其自由度分别为：

$\frac{1}{\sigma^2}Q_A$具有约束条件$\sum_{i=1}^{r}(\bar{\varepsilon}_i.-\bar{\varepsilon})=0$它的自由度为$r-1$；

$\frac{1}{\sigma^2}Q_B$具有约束条件$\sum_{j=1}^{s}(\bar{\varepsilon}.j-\bar{\varepsilon})=0$它的自由度为$s-1$；

$\frac{1}{\sigma^2}Q_E$具有约束条件$\sum_{i=1}^{r}(\bar{\varepsilon}_{ij}-\bar{\varepsilon}_i.-\bar{\varepsilon}._j+\bar{\varepsilon})=0(j=1,2,\cdots,s)$以

及$\sum_{j=1}^{s}(\varepsilon_{ij}-\bar{\varepsilon}_i.-\bar{\varepsilon}._j+\bar{\varepsilon})=0$，$i=1,2,\cdots,r$，

其中最后一个等式可由前面的$r+s-1$个获得，独立的约束条件有$r+s-1$个，故自由度为

$$rs-r-s+1=(r-1)(s-1), \frac{1}{\sigma^2} rs\bar{\varepsilon}^{-2}$$

的自由度为 1。由于右边各项自由度之和等于rs 。

因此，由柯赫伦因子分解定理知，$\frac{1}{\sigma^2}Q_A$服从自由度为$r-1$的χ^2 分布，$\frac{1}{\sigma^2}Q_B$服从自由度为$s-1$的χ^2 分布，$\frac{1}{\sigma^2}Q_E$服从自由度为$(r-1)(s-1)$的χ^2 分布。

且Q_A，Q_B和Q_E相互独立。

由F分布的定义

$$F_A = \frac{Q_A/\sigma^2(r-1)}{Q_E/\sigma^2(r-1)(s-1)} = \frac{\overline{Q}_A}{\overline{Q}_E} \sim F(r-1,(r-1)(s-1))$$

$$F_B = \frac{Q_B/\sigma^2(s-1)}{Q_E/\sigma^2(r-1)(s-1)} = \frac{\overline{Q}_B}{\overline{Q}_E} \sim F(s-1,(r-1)(s-1))$$

还可以证明，当H_{01}成立时，$F_A \sim F(r-1,(r-1)(s-1))$；

当H_{02}成立时，$F_B \sim F(r-1,(r-1)(s-1))$。

为了检验假设H_{01}，给定显著性水平α，查F 分布表可得

$F_\alpha(r-1,(r-1)(s-1))$的值，使得

$$P\{F_A \geq F_\alpha(r-1,(r-1)(s-1))\} = \alpha。$$

根据一次抽样后的样本值算得F_A，

若

$$F_A \geq F_\alpha(r-1,(r-1)(s-1)),$$

则拒绝H_{01}，即认为因素A对试验结果有显著影响。

若

$$F_A < F_\alpha(r-1,(r-1)(s-1)),$$

则接受H_{01}，即认为因素A对试验结果无显著影响。

同样，为了检验假设H_{02}，给定显著性水平α，查F分布表可得$F_{\alpha}(s-1,(r-1)(s-1))$的值，使得

$$P\{F_B \geq F_{\alpha}(s-1,(r-1)(s-1))\} = \alpha,$$

根据一次抽样后所得的样本值计算F_B的值，若

$$F_B \geq F_{\alpha}(s-1,(r-1)(s-1)),$$

则拒绝H_{02}，即认为因素B对试验结果无显著影响。

将整个分析过程列为因素方差分析表（见表 5.2.1），表中Q_T项的计算F_A和F_B值时并没时用到，它公起校核公式$Q_T = Q_A + Q_B + Q_E$的作用。

方差分析表中的离差平方和也可用下面一组公式来计算。令

$$T=\sum_{i=1}^{r}\sum_{j=1}^{s}X_{ij},\quad P=\frac{1}{rs}T^2\quad Q_I=\frac{1}{s}\sum_{i=1}^{r}\left(\sum_{j=1}^{s}X_{ij}\right)^2$$

$$Q_{II}=\frac{1}{r}\sum_{j=1}^{s}\left(\sum_{i=1}^{r}X_{ij}\right)^2\quad R=\sum_{i=1}^{r}\sum_{j=1}^{s}X_{ij}^2$$

从而

$$\begin{cases}Q_A=Q_{\mathrm{I}}-P,\\ Q_B=Q_{\mathrm{II}}-P,\\ Q_E=R-Q_{\mathrm{I}}-Q_{\mathrm{II}}+P\\ Q_{\mathrm{T}}=R-P.\end{cases}\qquad (5.2.20)$$

表5.2.1

方差来源	离差平方和	自由度	均方误差	F 值	显著性
因素A	$Q_A = s\sum_{i=1}^{r}(\bar{X}_i - \bar{X})^2$	$r-1$	$\bar{Q}_A = \frac{Q_A}{r-1}$	$F_A = \frac{\bar{Q}_A}{\bar{Q}_E}$	
因素B	$Q_B = r\sum_{j=1}^{s}(\bar{X}_{\cdot j} - \bar{X})^2$	$s-1$	$\bar{Q}_B = \frac{Q_B}{s-1}$	$F_B = \frac{\bar{Q}_B}{\bar{Q}_E}$	
误差	$Q_E = \sum_{i=1}^{r}\sum_{j=1}^{s}(\bar{X}_{ij} - \bar{X}_{i\cdot} - \bar{X}_{\cdot j} + \bar{X})^2$	$(r-1)(s-1)$	$\bar{Q}_E = \frac{Q_E}{(r-1)(s-1)}$		
总和	$Q_T = \sum_{i=1}^{r}\sum_{j=1}^{s}(\bar{X}_{ij} - \bar{X})^2$	$rs-1$			

例 5.2.1 为提高某种合金钢的强度，需要同时考察碳（C）及钛（Ti）的含量对强度的影响，以便选取合理的成分组合使强度达到最大。在试验中分别取因素A（C 含量%）3 个水平，因素 B（Ti 含量%）4 个水平，在组合水平 $(A_i, B_j), (i = 1,2,3, j = 1,2,3,4)$条件下各炼一炉钢，测得其强度数据见表 5.2.2。

试问：碳与钛的含量对合金钢的强度是否有显著影响$(\alpha = 0.01)$。

表5.2.2

B 水平 / A 水平	B_1 (3.3)	B_2 (3.4)	B_3 (3.5)	B_4 (3.6)
A_1 (0.03)	63.1	63.9	65.6	66.8
A_2 (0.04)	65.1	66.4	67.8	69.0
A_3 (0.05)	67.2	71.0	71.9	73.5

解　本例中$r=3, s=4, rs=12$，令X_{ij}为A与B的组合水平(A_i, B_j)下的试验结果。

$$Q_T = \sum_{i=1}^{3}\sum_{j=1}^{4}(X_{ij} - \overline{X})^2 = 113.29$$

$$Q_A = 4\sum_{i=1}^{3}(\overline{X}_{i}. - \overline{X})^2 = 74.91$$

$$Q_B = 3\sum_{j=1}^{3}(\overline{X}._{j} - \overline{X})^2 = 35.7$$

$$Q_E = Q_T - Q_A - Q_B = 3.21$$

方差分析表见表5. 2. 3

表5.2.3 方差分析表

方差来源	离差平方和	自由度	均方误差	F 值	显著性
因素A	74.91	2	37.46	70.02	* *
因素B	35.17	3	11.72	21.91	* *
误 差	3.21	6	0.535		
总 和	113.29	11			

查***F***分布表得：

$$F_{0.01}(2,6) = 10.9, F_{0.01}(3,6) = 9.78$$

由于

$$F_A = 70.02 > F_{0.01}(2,6) = 10.9, F_B = 21.91 > F_{0.01}(3,6) = 9.78$$

5.2 双因素方差分析

5.2.2 有交互作用

在前面模型中，由于只对A，B两个因素的每一种组合水平进行了一次试验，所以不能分析A，B两因素间是否存在交互作用的影响。下面讨论在每一种组合水平(A_i, B_j)下等重复试验情形的方差分析问题。

当两因素A与B间存在交互作用(interactions)时，为了考察交互作用就要做重复实验，如果每个处理的实验次数K都是相等的则称为平衡(balanced)实验.

二、两因素等重复试验的方差分析

数学模型

设有两个因素A和B，因素A有r个不同水平$A_1, A_2, \cdots, A_r$，因素B有s个不同水平$B_1, B_2, \cdots, B_s$，在每一种组合水平(A_i, B_j)下重复试验t次，测得试验数据为

$$X_{ijk}\,(i=1,2,\cdots,r, j=1,2,\cdots,s, k=1,2,\cdots,t)。$$

将它们列成表（表 5.12）。

表5.12

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

假定X_{ijk}服从正态分布

$$N(\mu_{ij}, \sigma^2)(i=1,2,\cdots,r, j=1,2,\cdots,s, k=1,2,\cdots,t),$$

且所有X_{ijk}相互独立，μ_{ij}可以表示为

$$\mu_{ij} = \mu + \alpha_i + \beta_j + \delta_{ij}, \qquad i=1,2,\cdots,s,$$

其中

$$\mu=\frac{1}{rs}\sum_{i=1}^{r}\sum_{j=1}^{s}\mu_{ij},\alpha_i=\frac{1}{s}\sum_{j=1}^{s}(\mu_{ij}-\mu)$$

$$\beta_j=\frac{1}{r}\sum_{i=1}^{r}(\mu_{ij}-\mu),\delta_{ij}=(\mu_{ij}-\mu-\alpha_i-\beta_j)$$

容易证明下列各式成立

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

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

从而得两因素等重复试验方差分析的数学模型为

$$X_{ijk} = \mu + \alpha_i + \beta_j + \delta_{ij} + \varepsilon_{ijk}, \qquad \varepsilon_{ijk} \sim N(0,\sigma^2) \tag{5.2.8}$$

$$i = 1,2,\cdots,r, j = 1,2,\cdots,s, k = 1,2,\cdots,t$$

其中诸ε_{ijk}相互独立，α_i称为因素A在水平A_i的效应，β_j称为因素B在水平B_j的效应，δ_{ij}称为因素A，B在组合水平(A_i, B_j)的交互作用效应。

因此，要判断因素A，B以及A与B交互作用$A\times B$的影响是否显著，分别等价于检验假设

$$H_{01}: \alpha_1 = \alpha_2 = \cdots = \alpha_r = 0 \leftrightarrow H_{11}: \alpha_1, \cdots, \alpha_r$$

中至少有一个不为0 。

$$H_{02}: \beta_1 = \beta_2 = \cdots = \beta_s = 0 \leftrightarrow H_{12}: \beta_1, \cdots, \beta_s$$

中至少有一个不为0。

$$H_{03}: \delta_{ij} = 0, i = 1,2,\cdots,r, j = 1,2,\cdots,s. \leftrightarrow H_{13}: \delta_{11}, \cdots, \delta_{rs}$$

中至少有一个不为0的问题.

为了导出上述三个假设检验的统计量，仍采取离差平方和分解的办法。

令

$$\overline{X} = \frac{1}{rst}\sum_{i=1}^{r}\sum_{j=1}^{s}\sum_{k=1}^{t} X_{ijk} \qquad \overline{X}_{ij\cdot} = \frac{1}{t}\sum_{k=1}^{t} X_{ijk}$$

$$\overline{X}_{i\cdot\cdot}=\frac{1}{st}\sum_{j=1}^{s}\sum_{k=1}^{t}X_{ijk}\qquad \overline{X}_{\cdot j\cdot}=\frac{1}{rt}\sum_{i=1}^{r}\sum_{k=1}^{t}X_{ijk}$$

于是有

$$Q_T=\sum_{i=1}^{r}\sum_{j=1}^{s}\sum_{k=1}^{t}(X_{ijk}-\overline{X})^2$$

$$=\sum_{i=1}^{r}\sum_{j=1}^{s}\sum_{k=1}^{t}[(\overline{X}_{i\cdot\cdot}-\overline{X})+\overline{X}_{\cdot j\cdot}-\overline{X})$$

$$+(\overline{X}_{ij\cdot}-\overline{X}_{i\cdot\cdot}-\overline{X}_{\cdot j\cdot}+\overline{X})+(X_{ijk}-\overline{X}_{ij\cdot})]^2$$

$$=\sum_{i=1}^{r}\sum_{j=1}^{s}\sum_{k=1}^{t}(\overline{X}_{i\cdot\cdot}-\overline{X})^2+\sum_{i=1}^{r}\sum_{j=1}^{s}\sum_{k=1}^{t}(\overline{X}_{\cdot j\cdot}-\overline{X})^2$$

$$+\sum_{i=1}^{r}\sum_{j=1}^{s}\sum_{k=1}^{t}(\overline{X}_{ij\cdot}-\overline{X}_{i\cdot\cdot}-\overline{X}_{\cdot j\cdot}+\overline{X})^2$$

$$+\sum_{i=1}^{r}\sum_{j=1}^{s}\sum_{k=1}^{t}(\overline{X}_{jki}-\overline{X}_{ij\cdot})^2 \qquad (5.2.10)$$

其中

$$
\begin{cases}
Q_A = \sum_{i=1}^{r}\sum_{j=1}^{s}\sum_{k=1}^{t}(\overline{X}_{i\cdot\cdot} - \overline{X})^2 = st\sum_{i=1}^{r}(\overline{X}_{i\cdot\cdot} - \overline{X})^2, \\
Q_B = \sum_{i=1}^{r}\sum_{j=1}^{s}\sum_{k=1}^{t}(\overline{X}_{\cdot j\cdot} - \overline{X})^2 = rt\sum_{j=1}^{s}(\overline{X}_{\cdot j\cdot} - \overline{X})^2, \\
Q_{A\times B} = \sum_{i=1}^{r}\sum_{j=1}^{s}\sum_{k=1}^{t}(\overline{X}_{ij\cdot} - \overline{X}_{i\cdot\cdot} - \overline{X}_{\cdot j\cdot} + \overline{X})^2 \\
\qquad = \sum_{i=1}^{r}\sum_{j=1}^{s}(\overline{X}_{ij\cdot} - \overline{X}_{i\cdot\cdot} - \overline{X}_{\cdot j\cdot} + \overline{X})^2, \\
Q_E = \sum_{i=1}^{r}\sum_{j=1}^{s}\sum_{k=1}^{t}(\overline{X}_{ijk} - \overline{X}_{ij\cdot})^2
\end{cases}
\tag{5.2.11}
$$

称Q_A为因素A引起的离差平方和，Q_B为因素 B 引起的离差平方和，$Q_{A\times B}$为因素A与B的交互作用$A\times B$引起的离差平方和，Q_E为误差平方和。

令

$$\bar{\varepsilon}=\frac{1}{rst}\sum_{i=1}^{r}\sum_{j=1}^{s}\sum_{k=1}^{t}\varepsilon_{ijk} \qquad \bar{\varepsilon}_{ij\cdot}=\frac{1}{t}\sum_{k=1}^{t}\varepsilon_{ijk}$$

$$\bar{\varepsilon}_{i\cdot\cdot}=\frac{1}{s}\sum_{j=1}^{s}\bar{\varepsilon}_{ij\cdot} \qquad \bar{\varepsilon}_{\cdot j\cdot}=\frac{1}{r}\sum_{i=1}^{r}\bar{\varepsilon}_{ij\cdot}$$

则可得

$$\left.\begin{aligned}
Q_A &= st\sum_{i=1}^{r}(\alpha_i+\bar{\varepsilon}_{i}..-\bar{\varepsilon})^2,\\
Q_B &= rt\sum_{j=1}^{s}(\beta_j+\bar{\varepsilon}._{j}.-\bar{\varepsilon})^2,\\
Q_{A\times B} &= t\sum_{i=1}^{r}\sum_{j=1}^{s}(\delta_{ij}+\bar{\varepsilon}_{ij}.-\bar{\varepsilon}_{i}..-\bar{\varepsilon}._{j}.+\bar{\varepsilon})^2,\\
Q_E &= \sum_{i=1}^{r}\sum_{j=1}^{s}\sum_{k=1}^{t}(\varepsilon_{ijk}-\bar{\varepsilon}_{ij}.)^2.
\end{aligned}\right\} \quad (5.2.12')$$

可算得它们的期望值分别为

$$
\left.\begin{aligned}
&E(Q_A)=(r-1)\sigma^2+st\sum_{i=1}^{r}\alpha_i^2,\\
&E(Q_B)=(s-1)\sigma^2+rt\sum_{j=1}^{s}\beta_j^2,\\
&E(Q_{A\times B})=(r-1)(s-1)\sigma^2+t\sum_{i=1}^{r}\sum_{j=1}^{s}\delta_{ij}^2,\\
&E(Q_E)=rs(t-1)\sigma^2.
\end{aligned}\right\}
\qquad (5.2.12)
$$

令 $\overline{Q}_A = \dfrac{Q_A}{r-1}$，$\overline{Q}_A$称为因素$A$引起的平均离差平方和，

$\overline{Q}_B = \dfrac{Q_B}{s-1}$，$\overline{Q}_B$称为因素$B$引起的平均离差平方和，

$Q_{A\times B} = \dfrac{Q_{A\times B}}{(r-1)(s-1)}$，$\overline{Q}_{A\times B}$称为因素$A$与$B$的交互作用引起的平均离差平方和，

$\overline{Q}_E = \dfrac{Q_E}{rs(t-1)}$，$\overline{Q}_E$称为平均离差平方和。

于是

$$E(\overline{Q}_A)=\sigma^2+\frac{st}{r-1}\sum_{i=1}^{r}\alpha_i^2,$$

$$E(\overline{Q}_B)=\sigma^2+\frac{rt}{s-1}\sum_{j=1}^{s}\beta_j^2,$$

$$E(\overline{Q}_{A\times B})=\sigma^2+\frac{t}{(r-1)(s-1)}\sum_{i=1}^{r}\sum_{j=1}^{s}\delta_{ij}^2,$$

$$E(\overline{Q}_E)=\sigma^2.$$

当H_{01}成立时，$E(\overline{Q}_A)=E(\overline{Q}_E)$，否则，$E(\overline{Q}_A)>E(\overline{Q}_E)$，

当H_{02}成立时，$E(\overline{Q}_B)=E(\overline{Q}_E)$，否则$E(\overline{Q}_B)>E(\overline{Q}_E)$；

当H_{03}成立时，$E(\overline{Q}_{A\times B})=E(\overline{Q}_E)$，否则有$E(\overline{Q}_{A\times B})>E(\overline{Q}_E)$，

令

$$F_A=\frac{\overline{Q}_A}{\overline{Q}_E},\qquad F_B=\frac{\overline{Q}_B}{\overline{Q}_E},$$

$$F_{A\times B}=\frac{\overline{Q}_{A\times B}}{\overline{Q}_E}.$$

则当H_{01}，H_{02}，H_{03}不成立时，F_A，F_B，$F_{A\times B}$都有偏大的趋势，因此F_A，F_B，$F_{A\times B}$可分别作为检验假设 H_{01}，H_{02}，H_{03}的统计量，下面导出F_A，F_B，及$F_{A\times B}$的概率分布。

在H_{01}，H_{02}和H_{03}都成立的条件下，由离差平方和分解公式（5.2.12）得

$$Q_T=\sum_{i=1}^{r}\sum_{j=1}^{s}\sum_{k=1}^{t}(X_{ijk}-\overline{X})^2=\sum_{i=1}^{r}\sum_{j=1}^{s}\sum_{k=1}^{t}(\varepsilon_{ijk}-\bar{\varepsilon})^2$$

$$=st\sum_{i=1}^{r}(\bar{\varepsilon}_{i}..-\bar{\varepsilon})^2+rt\sum_{j=1}^{s}(\bar{\varepsilon}._{j}.-\bar{\varepsilon})^2$$

$$+t\sum_{i=1}^{r}\sum_{j=1}^{s}(\bar{\varepsilon}_{ij}.-\bar{\varepsilon}_{i}..-\bar{\varepsilon}._{j}.+\bar{\varepsilon})^2+\sum_{i=1}^{r}\sum_{j=1}^{s}\sum_{k=1}^{t}(\varepsilon_{ijk}-\bar{\varepsilon}_{ij}.)^2,$$

上式两边同除以σ^2后，左边$\frac{1}{\sigma^2}Q_T$是自由度为$rst-1$的χ^2变量，而右边各项的自由度分别是：$\frac{1}{\sigma^2}Q_A$的自由度为$r-1$；$\frac{1}{\sigma^2}Q_B$的自由度为$s-1$；$\frac{1}{\sigma^2}Q_{A\times B}$的自由度为$(r-1)(s-1)$；$\frac{1}{\sigma^2}Q_E$的自由度为$rs(t-1)$。

由于右边各项自由度之和等于左边χ^2 变量的自由度，

故由柯赫伦因子分解定理知，

$$\frac{1}{\sigma^2}Q_A \sim \chi^2(r-1), \quad \frac{1}{\sigma^2}Q_B \sim \chi^2(s-1),$$

$$\frac{1}{\sigma^2}Q_{A\times B} \sim \chi^2(r-1)(s-1), \quad \frac{1}{\sigma^2}Q_E \sim \chi^2(rs(t-1)),$$

且它们相互独立。

从而由F 分布的定义知当H_{01}，H_{02} 和H_{03} 成立时，$F_A = \dfrac{\overline{Q}_A}{\overline{Q}_E}$

服从自由度为$(r-1, rs(t-1))$的F 分布；

$F_B = \dfrac{\overline{Q}_B}{\overline{Q}_E}$服从自由度为$(s-1, rs(t-1))$的$F$ 分布；

$F_{A\times B}=\dfrac{\overline{Q}_{A\times B}}{\overline{Q}_E}$服从自由度为$((r-1)(s-1),rs(t-1))$的 F 分布。

给定的显著性水平α，查F 分布表可得

$$F_\alpha[r-1,rs(t-1)]，F_\alpha[s-1,rs(t-1)]$$

和

$$F_\alpha[(r-1)(s-1),rs(t-1)]$$

的值，由一次抽样后所得的样本值算得F_A ,F_B 和$F_{A\times B}$的值。

若

$$F_A \geq F_\alpha[r-1, rs(t-1)],$$

则拒绝 H_{01}，即认为因素 A 对试验结果有显著影响，否则，接受 H_{01}，即认为因素 A 对试验结果无显著影响。

若

$$F_B \geq F_\alpha[s-1, rs(t-1)]$$

则拒绝 H_{02}，即认为因素 B 对试验结果有显著影响，否则，接受 H_{02}，即认为因素 B 对试验结果无显著影响。

若

$$F_{A\times B} \geq F_{\alpha}[(r-1)(s-1), rs(t-1)],$$

则拒绝H_{03}，即认为因素A与B的交互作用对试验结果有显著影响。否则，接受H_{03}，即认为因素A与B的交互作用对试验结果无显著影响。

将整个分析过程列成双因素方差分析表如下（见表5.2.5）.

表5. 2. 5

方差来源	离差平方和	自由度	平均离差平方和	F 值	显著性
因素 A	Q_A	$r-1$	$\overline{Q}_A = \dfrac{Q_A}{r-1}$	$F_A = \dfrac{\overline{Q}_A}{\overline{Q}_E}$	
因素 B	Q_B	$s-1$	$\overline{Q}_B = \dfrac{Q_B}{s-1}$	$F_B = \dfrac{\overline{Q}_B}{\overline{Q}_E}$	
交互作用 $A\times B$	$Q_{A\times B}$	$(r-1)(s-1)$	$\overline{Q}_{A\times B} = \dfrac{Q_{A\times B}}{(r-1)(s-1)}$	$F_{A\times B} = \dfrac{\overline{Q}_{A\times B}}{\overline{Q}_E}$	
误　差	Q_E	$rs(t-1)$	$\overline{Q}_E = \dfrac{Q_E}{rs(t-1)}$		
总　知	Q_T	$rst-1$			

表中离差平方和Q_A，Q_B，$Q_{A\times B}$，Q_E，Q_T 也可用如下公式计算。令

$$T=\sum_{i=1}^{r}\sum_{j=1}^{s}\sum_{k=1}^{t}X_{ijk}\,,\quad P=\frac{1}{rst}T^2$$

$$U=\frac{1}{st}\sum_{i=1}^{r}\left(\sum_{j=1}^{s}\sum_{k=1}^{t}X_{ijk}\right)^2,\quad V=\frac{1}{rt}\sum_{j=1}^{r}\left(\sum_{i=1}^{s}\sum_{k=1}^{t}X_{ijk}\right)^2$$

$$R=\frac{1}{t}\sum_{i=1}^{r}\sum_{j=1}^{s}\left(\sum_{k=1}^{t}X_{ijk}\right)^2,\quad W=\sum_{i=1}^{r}\sum_{j=1}^{s}\sum_{k=1}^{t}X_{ijk}^2\,,$$

则可得

$$Q_A = U - P,\quad Q_B = V - P,$$
$$Q_{A\times B} = R - U - V + P, \qquad (5.2.20)$$
$$Q_E = W - R,\quad Q_T = W - P。$$

例 5.2.2　考察合成纤维中对纤维弹性有影响的两个因素，收缩率A和总拉伸倍数B。A和B各取 4 种水平，每种组合水平重复试验两次，得数据见表

表5.2.6

因素 A / 因素 B	460(B_1)	520(B_2)	580(B_3)	640(B_4)
0(A_1)	71, 73	72, 73	75, 73	77, 75
4(A_2)	73, 75	76, 74	78, 77	74, 74
8(A_3)	76, 73	79, 77	74, 75	74, 73
12(A_4)	75, 73	73, 72	70, 71	69, 69

试问收缩率和总位伸倍数分别对纤维弹性有无显著影响？收缩率与总拉伸倍数之间的交互作用是否影响显著？$(\alpha = 0.05)$

解　依题意有$r=s=4,t=2$。F_A，F_B和$F_{A\times B}$值的计算按如下双因素方差分析表进行（表 4. 2. 7）。

由$\alpha=0.05$，查F 分布表

$$F_{0.05}(3,16)=3.24，\quad F_{0.05}(9,16)=2.54$$

比较知

$$F_A=17.5>3.24=F_{0.05}(3,16),$$

$$F_B=2.1>3.24=F_{0.05}(3,16),$$

$$F_{A\times B}=6.6>2.54=F_{0.05}(9,16),$$

故合成纤维收缩率对弹性有显著影响，总拉伸倍数对弹性无显著影响，而收缩率和总拉伸倍数的交互作用对弹性有显著影响。

表 5.2.7

方差来源	离差平方和	自由度	均方误差	F 值	显著性
收缩率 A	70.594	3	$S_A^2 = 23.531$	$F_A = 17.5$	*
总拉伸倍数 B	8.594	3	$S_B^2 = 2.865$	$F_B = 2.1$	
交互作用 $A\times B$	79.531	9	$S_{A\times B}^2 = 8.837$	$F_{A\times B} = 6.6$	*
误　差	21.500	16	$S_E^2 = 1.344$		
总　和	180.219	31			

谢 谢!