第5.2节　两因素方差分析

一、双因素非重复试验的方差分析

二、双因素等重复试验的方差分析

一、两因素非重复试验的方差分析

1、两因素试验

设有两个因素A, B,因素A 有r 个不同的水平;B有s个不同水平，A, B 的每一种组合水平(A_i, B_j)下做一次试验，试验结果为X_{ij},所有X_{ij} 相互独立，所有试验结果为

因素B / 因素A	B_1	B_2	$\cdots$	B_s
A_1	X_{11}	X_{12}	$\cdots$	X_{1s}
A_2	X_{21}	X_{22}	$\cdots$	X_{2s}
$\vdots$	$\vdots$	$\vdots$		$\vdots$
A_r	X_{r1}	X_{r2}	$\cdots$	X_{rs}

为了方便计算，记$\overline{X}_{\cdot j}=\dfrac{1}{r}\sum_{i=1}^{r}X_{ij},\overline{X}_{i\cdot}=\dfrac{1}{s}\sum_{j=1}^{s}X_{ij},$

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

2、两因素非重复试验

对两个因素的每一组合只做一次试验, 称其为**两因素非重复试验**.

3、数学模型

假设 $X_{ij} \sim N(\mu_{ij}, \sigma^2), i = 1, \cdots, r, j = 1, \cdots, s.$ 各 X_{ij} 独立, μ_{ij}, σ^2 均为未知参数.

$$\left.\begin{array}{l} X_{ij} = \mu_{ij} + \varepsilon_{ij}, \\ i = 1,2,\cdots,r, j = 1,2,\cdots,s, \\ \varepsilon_{ij} \sim N(0,\sigma^2), \text{各} \varepsilon_{ij} \text{独立}, \end{array}\right\}$$

为了应用方差分析此模型，需要做如下变换

设 $\boxed{\mu = \frac{1}{rs}\sum_{i=1}^{r}\sum_{j=1}^{s}\mu_{ij}}$ —总平均

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

$\boxed{\alpha_i = \mu_{i\bullet} - \mu, \quad i = 1,\cdots,r}$ —水平A_i 的效应，表示A_i 在总体平均数上引起的偏差

$\boxed{\beta_j = \mu_{\bullet j} - \mu, \quad j = 1,\cdots,s}$ —水平B_j 的效应，表示B_j 在总体平均数上引起的偏差

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

$$\mu_{ij} = \mu + \alpha_i + \beta_j + \boxed{(\mu_{ij} - \alpha_i - \beta_j - \mu)}$$

$$= \mu + \alpha_i + \beta_j + \delta_{ij},$$

因素A, B在组合水平(A_i, B_j)交互效应

其中$\delta_{ij} = \mu_{ij} - \alpha_i - \beta_j - \mu, \sum_{i=1}^{r} \delta_{ij} = 0, j = 1, \cdots, s$

$\sum_{j=1}^{s} \delta_{ij} = 0, i = 1, \cdots, r$.又由于只做一次试验，没有

交互效应，因而$\delta_{ij} = 0$.因而原模型可以转化为

$$\left.\begin{array}{l} X_{ij}=\mu+\alpha_i+\beta_j+\varepsilon_{ij}, \\ \varepsilon_{ij}\sim N(0,\sigma^2),\varepsilon_{ij}\text{相互独立}, \\ i=1,2,\cdots,r, j=1,2,\cdots,s, \\ \sum_{i=1}^{r}\alpha_i=0,\sum_{j=1}^{s}\beta_j=0. \end{array}\right\}$$

称其为两因素非重复试验方差分析的数学模型

要推断因素A的影响是否显著，就等价于假设检验

$$H_{01}:\alpha_1=\cdots=\alpha_r=0,\leftrightarrow H_{11}:\alpha_1,\cdots,\alpha_r\text{不全为零}.$$

要推断因素B的影响是否显著，就等价于假设检验

$$H_{02}:\beta_1=\cdots=\beta_s=0,\leftrightarrow H_{12}:\beta_1,\cdots,\beta_s\text{不全为零}.$$

3. 离差平方和分解

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

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

$$\boxed{Q_T = \sum_{i=1}^{r}\sum_{j=1}^{s}(X_{ij} - \bar{X})^2}$$ 总离差平方和(总变差)

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

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

$$+\sum_{i=1}^{r}\sum_{j=1}^{s}[(\bar{X}_{\bullet j} - \bar{X})]^2$$

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

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

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

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

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

交叉项等于零后面给出证明

$$Q_A = s\sum_{i=1}^{r}(\bar{X}_{i\bullet} - \bar{X})^2 \qquad Q_B = r\sum_{j=1}^{s}(\bar{X}_{\bullet j} - \bar{X})^2$$

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

$$Q_T = \boxed{Q_E} \boxed{+Q_A} \boxed{+Q_B}$$

随机误差平方和

因素A引起的离差平方和

因素B引起的离差平方和

上式成立是由于三个交叉项相乘都等于零

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

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

$$=\sum_{i=1}^{r}[s\bar{X}_{i\cdot}-s\bar{X}_{i\cdot}-s\bar{X}+s\bar{X}]\left(\bar{X}_{i\cdot}-\bar{X}\right)=0$$

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

$$=\sum_{j=1}^{s}[r\bar{X}_{\cdot j}-r\bar{X}-r\bar{X}_{\cdot j}+r\bar{X}]\left(\bar{X}_{\cdot j}-\bar{X}\right)=0$$

$$\sum_{i=1}^{r}\sum_{j=1}^{s}\left(\bar{X}_{i.}-\bar{X}\right)\left(\bar{X}_{.j}-\bar{X}\right)$$

$$=\left[\sum_{j=1}^{s}\left(\bar{X}_{.j}-\bar{X}\right)\right]\left[\sum_{i=1}^{r}\left(\bar{X}_{i.}-\bar{X}\right)\right]$$

$$=(s\bar{X}-s\bar{X})(r\bar{X}-r\bar{X})=0$$

3. 离差平方和的统计特性

令 $\bar{\varepsilon}_{i\cdot}=\dfrac{1}{s}\sum_{j=1}^{s}\varepsilon_{ij},\ i=1,\cdots,r,\quad \bar{\varepsilon}_{\cdot j}=\dfrac{1}{r}\sum_{i=1}^{r}\varepsilon_{ij},\ j=1,\cdots,s,$

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

则

$$Q_A=s\sum_{i=1}^{r}\left(\frac{1}{s}\sum_{j=1}^{s}X_{ij}-\frac{1}{rs}\sum_{i=1}^{r}\sum_{j=1}^{s}X_{ij}\right)^2=s\sum_{i=1}^{r}\left[\frac{1}{s}\sum_{j=1}^{s}\left(X_{ij}-\frac{1}{r}\sum_{i=1}^{r}X_{ij}\right)\right]^2$$

$$=s\sum_{i=1}^{r}\left\{\frac{1}{s}\sum_{j=1}^{s}\left[(\mu+\alpha_i+\beta_j+\varepsilon_{ij})-\frac{1}{r}\sum_{i=1}^{r}(\mu+\alpha_i+\beta_j+\varepsilon_{ij})\right]\right\}^2$$

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

同理

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

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

又因为

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

$$\bar{\varepsilon} \sim N(0, \frac{\sigma^2}{sr})$$

因而

$$
\begin{aligned}
E\left(Q_A\right) &= sE\left[\sum_{i=1}^{r}\left(\alpha_i+\overline{\varepsilon}_{i.}-\overline{\varepsilon}\right)^2\right] \\
&= s\sum_{i=1}^{r}\alpha_i^2+sE\left[\sum_{i=1}^{r}\left(\overline{\varepsilon}_{i.}-\overline{\varepsilon}\right)^2\right]+2s\sum_{i=1}^{r}\alpha_i E\left(\overline{\varepsilon}_{i.}-\overline{\varepsilon}\right) \\
&= s\sum_{i=1}^{r}\alpha_i^2+sE[\sum_{i=1}^{r}\overline{\varepsilon}_{i.}^2-r\overline{\varepsilon}^2]+2s\sum_{i=1}^{r}\alpha_i[E\overline{\varepsilon}_{i.}-E\overline{\varepsilon}] \\
&= s\sum_{i=1}^{r}\alpha_i^2+\left(r-1\right)\sigma^2+0 \\
&= s\sum_{i=1}^{r}\alpha_i^2+\left(r-1\right)\sigma^2
\end{aligned}
$$

同理

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

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

令 $\overline{Q}_A = \dfrac{1}{r-1}Q_A,\ \overline{Q}_B = \dfrac{1}{s-1}Q_B,\ \overline{Q}_E = \dfrac{1}{(r-1)(s-1)}Q_E,$

则

$$E\overline{Q}_A = \sigma^2 + \frac{s}{r-1}\sum_{i=1}^{r}\alpha_i^2,\ E\overline{Q}_B = \frac{r}{s-1}\sum_{i=1}^{r}\beta_i^2 + \sigma^2,$$

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

由于H_{01}成立时，$E(\bar{Q}_A)=E\bar{Q}_E$,否则$E(\bar{Q}_A)>E\bar{Q}_E$;

由于H_{02}成立时，$E(\bar{Q}_B)=E\bar{Q}_E$,否则$E(\bar{Q}_B)>E\bar{Q}_E$,

因此 构造统计量

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

4. 统计量的分布

由于H_{01}, H_{02}成立时，$\alpha_i=\beta_j=0(i=1,\cdots,r,\ j=1,\cdots,s)$

因而$X_{ij}=\mu+\varepsilon_{ij}$,则离差平方和可以改写为

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

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

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

又由于$\dfrac{\varepsilon_{ij}}{\sigma} \sim N(0,1)$,由定理1.12可知，

$$\frac{Q_T}{\sigma^2} = \sum_{i=1}^{r}\sum_{j=1}^{s}\left(\frac{\varepsilon_{ij} - \bar{\varepsilon}}{\sigma}\right)^2 \sim \chi^2(rs-1).$$

$$\frac{Q_A}{\sigma^2} = \sum_{i=1}^{r}\left(\frac{\bar{\varepsilon}_{i\cdot} - \bar{\varepsilon}}{\sigma/\sqrt{s}}\right)^2 \sim \chi^2(r-1).$$

$$\frac{Q_B}{\sigma^2}=\sum_{i=1}^{s}\left(\frac{\overline{\varepsilon}_{\cdot j}-\overline{\varepsilon}}{\sigma/\sqrt{r}}\right)^2\sim\chi^2(s-1).$$

而$\dfrac{1}{\sigma^2}Q_E$具有约束$\sum\limits_{i=1}^{r}(\varepsilon_{ij}-\overline{\varepsilon}_{i\cdot}-\overline{\varepsilon}_{\cdot j}+\overline{\varepsilon})=0(j=1,\cdots,s)$

以及约束$\sum\limits_{j=1}^{s}(\varepsilon_{ij}-\overline{\varepsilon}_{i\cdot}-\overline{\varepsilon}_{\cdot j}+\overline{\varepsilon})=0(i=1,\cdots,r)$,而最后

一个约束可以由前$s+r-1$得到，因而其独立约束条件共$s+r-1$.

显然，离差平方和公式的左右两边自由度满足：

$$rs-1=(r-1)+(s-1)+(rs-r-s+1)$$

由柯赫伦因子分解定理(p16定理1.7)可知:

$$\frac{1}{\sigma^2} Q_E \sim \chi^2(rs - r - s + 1)$$

因而

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

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

5. 方差分析对应的拒绝域

在给定显著性水平α下，因素A 对试验结果有显著影响的拒绝域为

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

在给定显著性水平α下，因素B 对试验结果有显著影响的拒绝域为

$$W_{\mathrm{B}} = \{F_B \mid F_B \geq F_\alpha(s-1,(r-1)(s-1))\}$$

表5.9　双因素非重复试验的方差分析表

方差来源	平方和	自由度	均　方	F　比
因素A	Q_A	$r-1$	$\bar{Q}_A=\dfrac{Q_A}{r-1}$	$F_A=\dfrac{\bar{Q}_A}{\bar{Q}_E}$
因素B	Q_B	$s-1$	$\bar{Q}_B=\dfrac{Q_B}{s-1}$	$F_B=\dfrac{\bar{Q}_B}{\bar{Q}_E}$
误　差	Q_E	$(r-1)(s-1)$	$\bar{Q}_E=\dfrac{Q_E}{(r-1)(s-1)}$	
总　和	Q_T	$rs-1$		

为了计算方便，通常可以采用如下公式：令

$$T=\sum_{i=1}^{r}\sum_{j=1}^{s}X_{ij},\quad P=\frac{1}{rs}T^2,\quad R=\sum_{i=1}^{r}\sum_{j=1}^{s}X_{ij}^2,$$

$$Q_I=\frac{1}{s}\sum_{i=1}^{r}\Big(\sum_{j=1}^{s}X_{ij}\Big)^2,\quad Q_{II}=\frac{1}{r}\sum_{j=1}^{s}\Big(\sum_{i=1}^{r}X_{ij}\Big)^2,$$

则
$$\begin{cases} Q_A=Q_I-P, \\ Q_B=Q_{II}-P, \\ Q_E=R-Q_I-Q_{II}+P, \\ Q_T=R-P \end{cases}$$

例1(p165例5.5)为了提高某种合金钢的强度，需要同时考察炭C以及钛Ti的含量对强度的影响，以便选取合理的成份组合使得强度达到最大，在试验中分别取因素A(C的含量)3个水平，因素B(Ti的含量)4个水平，在组合$(A_i, B_j)(i=1,2,3, j=1,2,3,4)$条件下各炼一炉测得的强度为:

A水平 \ B水平	B_1	B_2	B_3	B_2
A_1	63.1	63.9	65.6	66.8
A_2	65.1	66.4	67.8	69.0
A_3	67.2	71.0	71.9	73.5

试问：炭与钛的含量对合金钢的强度是否有显著影响(α=0.01).

解 $r=3, s=4, rs=12$,经计算

$$Q_T=113.29,\ \ Q_A=74.91,\ \ Q_B=35.7,\ \ Q_E=3.21$$

$$F_A=\frac{\overline{Q}_A}{\overline{Q}_E}=70.02>F_{0.01}(2,6)=10.9$$

$$F_B=\frac{\overline{Q}_B}{\overline{Q}_B}=21.91>F_{0.01}(3,6)=9.78$$

因而炭与钛的含量对合金钢的强度是有显著影响.

二、两因素等重复试验的方差分析

因素$A: A_1, A_2, \cdots, A_r$, 因素$B: B_1, B_2, \cdots, B_s$,每一个组合水平(A_i, B_j)下重复试验t次，测得的数据为X_{ijk},如表.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	$X_{r11}, X_{r12}, \cdots, X_{r1t}$	$X_{r21}, X_{r22}, \cdots, X_{r2t}$	$\cdots$	$X_{rs1}, X_{rs2}, \cdots, X_{rst}$

1. 数学模型

假设

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

各X_{ijk}独立，μ_{ij}, σ^2均为未知参数.

$$\left.\begin{array}{l} X_{ijk} = \mu_{ij} + \varepsilon_{ijk}, \\ \varepsilon_{ijk} \sim N(0, \sigma^2), \text{各}\varepsilon_{ijk}\text{独立}, \\ i = 1,2,\cdots,r, j = 1,2,\cdots,s, \\ k = 1,2,\cdots,t. \end{array}\right\}$$

设 $$\boxed{\mu = \frac{1}{rs}\sum_{i=1}^{r}\sum_{j=1}^{s}\mu_{ij}}$$ —总平均

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

$$\boxed{\alpha_i = \mu_{i\bullet} - \mu, \quad i = 1,\cdots,r}$$

—水平A_i的效应，表示A_i在总体平均数上引起的偏差

$$\boxed{\beta_j = \mu_{\bullet j} - \mu, \quad j = 1,\cdots,s}$$

—水平B_j的效应，表示B_j在总体平均数上引起的偏差

$$\delta_{ij} = \boxed{\mu_{ij} - \mu - \alpha_i - \beta_j,}$$ —组合水平(A_i, B_j)的交互作用效应

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

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

证明

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

$$= \frac{1}{s} \sum_{i=1}^{r} \sum_{j=1}^{s} \mu_{ij} - \frac{1}{s} \sum_{i=1}^{r} \sum_{j=1}^{s} \mu = \frac{1}{s} rs\mu - \frac{1}{s} rs\mu = 0$$

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

$$= \frac{1}{r}\sum_{j=1}^{s}\sum_{i=1}^{r}\mu_{ij} - \frac{1}{r}\sum_{j=1}^{s}\sum_{i=1}^{r}\mu = \frac{1}{r}sr\mu - \frac{1}{r}sr\mu = 0$$

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

$$= \sum_{i=1}^{r}\mu_{ij} - \sum_{i=1}^{r}\mu - \sum_{i=1}^{r}\alpha_i - \sum_{i=1}^{r}\beta_j$$

$$= \sum_{i=1}^{r}(\mu_{ij}-\mu) - r\beta_j = 0$$

$$\sum_{j=1}^{s}\delta_{ij}=\sum_{j=1}^{s}(\mu_{ij}-\mu-\alpha_i-\beta_j)$$

$$=\sum_{j=1}^{s}\mu_{ij}-\sum_{j=1}^{s}\mu-\sum_{j=1}^{s}\alpha_i-\sum_{j=1}^{s}\beta_j$$

$$=\sum_{j=1}^{s}(\mu_{ij}-\mu)-s\alpha_i=0$$

$$\left.\begin{array}{l} X_{ijk}=\mu+\alpha_i+\beta_j+\delta_{ij}+\varepsilon_{ijk}, \\ \varepsilon_{ijk}\sim N(0,\sigma^2),\text{各}\varepsilon_{ijk}\text{相互独立}, \\ i=1,2,\cdots,r,\ j=1,2,\cdots,s,\ k=1,2,\cdots,t, \\ \sum_{i=1}^{r}\alpha_i=0,\sum_{j=1}^{s}\beta_j=0,\sum_{i=1}^{r}\delta_{ij}=0,\sum_{j=1}^{s}\delta_{ij}=0. \end{array}\right\}$$

称其为两因素等重复试验方差分析的数学模型

判断因素以及因素的交互作用对试验结果是否有显著影响等价于检验假设：

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

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

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

2.分解离差平方和

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

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

$$\boxed{Q_T = \sum_{i=1}^{r}\sum_{j=1}^{s}\sum_{k=1}^{t}(X_{ijk} - \overline{X})^2}$$ 总离差平方和(总变差)

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

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

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

$$= \sum_{i=1}^{r}\sum_{j=1}^{s}\sum_{k=1}^{t}(X_{ijk} - \overline{X}_{ij\bullet})^2 + st\sum_{i=1}^{r}(\overline{X}_{i\bullet\bullet} - \overline{X})^2 + rt\sum_{j=1}^{s}(\overline{X}_{\bullet j\bullet} - \overline{X})^2 + t\sum_{i=1}^{r}\sum_{j=1}^{s}(\overline{X}_{ij\bullet} - \overline{X}_{i\bullet\bullet} - \overline{X}_{\bullet j\bullet} + \overline{X})^2$$

$$Q_T = \boxed{Q_E} + \boxed{Q_A} + \boxed{Q_B} + \boxed{Q_{A\times B}}$$

误　差平方和

因素 A 的效应平方和

因素 B 的效应平方和

因素A,B的交互效应平方和

这里仅证明两个交叉项相乘等于零，其余类似可证）

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

$$=\sum_{j=1}^{s}\sum_{k=1}^{t}(\bar{X}_{\cdot j\cdot}-\bar{X})\left[\sum_{i=1}^{r}(\bar{X}_{i\cdot\cdot}-\bar{X})\right]$$

$$=\sum_{j=1}^{s}\sum_{k=1}^{t}(\bar{X}_{\cdot j\cdot}-\bar{X})\left[\sum_{i=1}^{r}\left(\frac{1}{st}\sum_{j=1}^{s}\sum_{k=1}^{t}X_{ijk}-\bar{X}\right)\right]$$

$$=\sum_{j=1}^{s}\sum_{k=1}^{t}(\bar{X}_{\cdot j\cdot}-\bar{X})\left[\frac{1}{st}\sum_{i=1}^{r}\sum_{j=1}^{s}\sum_{k=1}^{t}X_{ijk}-\sum_{i=1}^{r}\bar{X}\right]$$

$$=\sum_{j=1}^{s}\sum_{k=1}^{t}(\bar{X}_{\cdot j\cdot}-\bar{X})\left[r\frac{1}{rst}\sum_{i=1}^{r}\sum_{j=1}^{s}\sum_{k=1}^{t}X_{ijk}-r\bar{X}\right]=0$$

$$
\begin{aligned}
&\sum_{i=1}^{r}\sum_{j=1}^{s}\sum_{k=1}^{t}(\bar{X}_{i\cdot\cdot}-\bar{X})(\overline{X}_{ij\cdot}-\bar{X}_{i\cdot\cdot}-\bar{X}_{\cdot j\cdot}+\bar{X}) \\
&=\sum_{i=1}^{r}(\bar{X}_{i\cdot\cdot}-\bar{X})\left[\sum_{j=1}^{s}\sum_{k=1}^{t}(\overline{X}_{ij\cdot}-\bar{X}_{i\cdot\cdot}-\bar{X}_{\cdot j\cdot}+\bar{X})\right] \\
&=\sum_{i=1}^{r}(\bar{X}_{i\cdot\cdot}-\bar{X})\left[\sum_{j=1}^{s}\sum_{k=1}^{t}\overline{X}_{ij\cdot}-\sum_{j=1}^{s}\sum_{k=1}^{t}\bar{X}_{i\cdot\cdot}-\sum_{j=1}^{s}\sum_{k=1}^{t}\bar{X}_{\cdot j\cdot}+\sum_{j=1}^{s}\sum_{k=1}^{t}\bar{X}\right] \\
&=\sum_{i=1}^{r}(\bar{X}_{i\cdot\cdot}-\bar{X})\left[\sum_{j=1}^{s}t\overline{X}_{ij\cdot}-st\bar{X}_{i\cdot\cdot}-\sum_{j=1}^{s}t\bar{X}_{\cdot j\cdot}+\sum_{j=1}^{s}\sum_{k=1}^{t}\bar{X}\right] \\
&=\sum_{i=1}^{r}(\bar{X}_{i\cdot\cdot}-\bar{X})\left[st\bar{X}_{i\cdot\cdot}-st\bar{X}_{i\cdot\cdot}-st\bar{X}+st\bar{X}\right]=0
\end{aligned}
$$

3. 离差平方和的统计特性

令 $\bar{\varepsilon}=\dfrac{1}{rst}\sum\limits_{i=1}^{r}\sum\limits_{j=1}^{s}\sum\limits_{k=1}^{t}\varepsilon_{ijk}$, $\quad \bar{\varepsilon}_{ij\cdot}=\dfrac{1}{t}\sum\limits_{k=1}^{t}\varepsilon_{ijk}$,

$\bar{\varepsilon}_{i\cdot\cdot}=\dfrac{1}{s}\sum\limits_{j=1}^{s}\bar{\varepsilon}_{ij\cdot},\quad i=1,\cdots,r,\quad \bar{\varepsilon}_{\cdot j\cdot}=\dfrac{1}{r}\sum\limits_{i=1}^{r}\bar{\varepsilon}_{ij\cdot},\quad j=1,\cdots,s,$

易得

$$\left.\begin{aligned}
Q_A &= st\sum_{i=1}^{r}(\alpha_i+\bar{\varepsilon}_{i\cdot\cdot}-\bar{\varepsilon})^2\\
Q_B &= rt\sum_{i=1}^{r}(\beta_j+\bar{\varepsilon}_{\cdot j\cdot}-\bar{\varepsilon})^2\\
Q_{A\times B} &= t\sum_{i=1}^{r}\sum_{j=1}^{s}(\delta_{ij}+\bar{\varepsilon}_{ij\cdot}-\bar{\varepsilon}_{i\cdot\cdot}-\bar{\varepsilon}_{\cdot j\cdot}+\bar{\varepsilon})^2\\
Q_E &= \sum_{i=1}^{r}\sum_{j=1}^{s}\sum_{k=1}^{t}(\varepsilon_{ijk}-\bar{\varepsilon}_{ij\cdot})^2
\end{aligned}\right\}$$

这里仅给出 Q_A 的推导

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

$$= st\sum_{i=1}^{r}\left[\frac{1}{st}\sum_{j=1}^{s}\sum_{k=1}^{t}\left(X_{ijk} - \frac{1}{r}\sum_{i=1}^{r}X_{ijk}\right)\right]^2$$

$$= st\sum_{i=1}^{r}\left\{\frac{1}{st}\sum_{j=1}^{s}\sum_{k=1}^{t}\left[(\mu+\alpha_i+\beta_j+\delta_{ij}+\varepsilon_{ijk}) - \frac{1}{r}\sum_{i=1}^{r}(\mu+\alpha_i+\beta_j+\delta_{ij}+\varepsilon_{ijk})\right]\right\}^2$$

$$= st\sum_{i=1}^{r}\left\{\frac{1}{st}\sum_{j=1}^{s}\sum_{k=1}^{t}\left[\alpha_i+\beta_j+\delta_{ij}+\varepsilon_{ijk}-\beta_j-\frac{1}{r}\sum_{i=1}^{r}\varepsilon_{ijk}\right]\right\}^2$$

$$= st\sum_{i=1}^{r}\left\{\frac{1}{st}\sum_{j=1}^{s}\sum_{k=1}^{t}\left[\alpha_i+\delta_{ij}+\varepsilon_{ijk}-\frac{1}{r}\sum_{i=1}^{r}\varepsilon_{ijk}\right]\right\}^2$$

$$= st\sum_{i=1}^{r}(\alpha_i+\bar{\varepsilon}_{i\cdot\cdot}-\bar{\varepsilon})^2$$

又因为

$$\varepsilon_{ijk}\sim N(0,\sigma^2),\quad \bar{\varepsilon}_{i\cdot\cdot}\sim N(0,\frac{\sigma^2}{st}),\quad \bar{\varepsilon}_{\cdot j\cdot}\sim N(0,\frac{\sigma^2}{rt})$$

$$\bar{\varepsilon}_{ij\cdot}\sim N(0,\frac{\sigma^2}{t}),\quad \bar{\varepsilon}\sim N(0,\frac{\sigma^2}{srt}),$$

因此

$$EQ_A = E\left[st\sum_{i=1}^{r}(\alpha_i + \overline{\varepsilon}_{i\cdot\cdot} - \overline{\varepsilon})^2 \right]$$

$$= stE\left[\sum_{i=1}^{r}\alpha_i^2 + 2\sum_{i=1}^{r}\alpha_i(\overline{\varepsilon}_{i\cdot\cdot} - \overline{\varepsilon}) + \sum_{i=1}^{r}(\overline{\varepsilon}_{i\cdot\cdot} - \overline{\varepsilon})^2 \right]$$

$$= st\sum_{i=1}^{r}\alpha_i^2 + 2st\alpha_i\sum_{i=1}^{r}E(\overline{\varepsilon}_{i\cdot\cdot} - \overline{\varepsilon}) + stE\sum_{i=1}^{r}(\overline{\varepsilon}_{i\cdot\cdot} - \overline{\varepsilon})^2$$

$$= st\sum_{i=1}^{r}\alpha_i^2 + st\left[\sum_{i=1}^{r}E(\overline{\varepsilon}_{i\cdot\cdot}^2) - rE(\overline{\varepsilon}^2) \right]$$

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

同理 $E(Q_B)=r\sum_{i=1}^{r}\beta_i^2+(s-1)\sigma^2 \quad E(Q_E)=rs(t-1)\sigma^2$

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

令

$$\bar{Q}_A=\frac{Q_A}{r-1},\ \bar{Q}_B=\frac{Q_B}{s-1},\ \bar{Q}_{A\times B}=\frac{Q_{A\times B}}{(r-1)(s-1)},$$

$$\bar{Q}_E=\frac{Q_E}{rs(t-1)}$$

则

$$E\bar{Q}_A=\sigma^2+\frac{st}{r-1}\sum_{i=1}^{r}\alpha_i^2,\ E\bar{Q}_B=\frac{rt}{s-1}\sum_{i=1}^{r}\beta_i^2+\sigma^2,$$

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

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

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

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

因此 构造统计量

$$F_A=\frac{\overline{Q}_A}{\overline{Q}_E},\qquad F_B=\frac{\overline{Q}_B}{\overline{Q}_E},\qquad F_{A\times B}=\frac{\overline{Q}_{A\times B}}{\overline{Q}_E}$$

4. 统计量的分布

由于H_{01},H_{02},H_{03}成立时，$\alpha_i=\beta_j=\delta_{ij}=0(i=1,\cdots,r,$ $j=1,\cdots,s)$，因而$X_{ijk}=\mu+\varepsilon_{ijk}$，则离差平方和可以改写为

$$Q_A = st\sum_{i=1}^{r}(\overline{\varepsilon}_{i\cdot\cdot}-\overline{\varepsilon})^2, Q_B = rt\sum_{j=1}^{s}(\overline{\varepsilon}_{\cdot j\cdot}-\overline{\varepsilon})^2$$

$$Q_{A\times B} = t\sum_{i=1}^{r}\sum_{j=1}^{s}[\overline{\varepsilon}_{ij\cdot}-\overline{\varepsilon}_{i\cdot\cdot}-\overline{\varepsilon}_{\cdot j\cdot}+\overline{\varepsilon}]^2$$

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

又由于$\dfrac{\varepsilon_{ijk}}{\sigma} \sim N(0,1)$,由定理1.12可知,

$$\frac{Q_T}{\sigma^2} = \frac{1}{\sigma^2}\sum_{i=1}^{r}\sum_{j=1}^{s}\sum_{k=1}^{t}(\varepsilon_{ijk}-\overline{\varepsilon})^2 \sim \chi^2(rst-1).$$

$\dfrac{Q_A}{\sigma^2}$的自由度为$r-1$, $\dfrac{Q_B}{\sigma^2}$的自由度为$s-1$, $\dfrac{Q_{A\times B}}{\sigma^2}$的自由度为$(s-1)(r-1)$,, $\dfrac{Q_E}{\sigma^2}$的自由度为$rs(t-1)$,

显然，离差平方和公式的左右两边自由度满足：

$$rst-1=(r-1)+(s-1)+(rs-r-s+1)+rs(t-1)$$

由柯赫伦因子分解定理(p26定理1.7)可知:

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

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

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

$$\frac{1}{\sigma^2}Q_E \sim \chi^2(rst-rs)$$

因而

$$F_A = \frac{\dfrac{Q_A}{\sigma^2(r-1)}}{\dfrac{Q_E}{\sigma^2 rs(t-1)}} = \frac{\bar{Q}_A}{\bar{Q}_E} \sim F(r-1, rs(t-1))$$

$$F_B = \frac{\dfrac{Q_B}{\sigma^2(s-1)}}{\dfrac{Q_E}{\sigma^2 rs(t-1)}} = \frac{\bar{Q}_B}{\bar{Q}_E} \sim F(s-1, rs(t-1))$$

$$F_{A\times B} = \frac{\dfrac{Q_{A\times B}}{\sigma^2(s-1)(r-1)}}{\dfrac{Q_E}{\sigma^2 rs(t-1)}} = \frac{\bar{Q}_{A\times B}}{\bar{Q}_E} \sim F((s-1)(r-1), rs(t-1))$$

5. 方差分析对应的拒绝域

在给定显著性水平α下，因素A对试验结果有显著影响的拒绝域为

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

在给定显著性水平α下，因素B对试验结果有显著影响的拒绝域为

$$W_{\mathrm{B}} = \{F_B \mid F_B \geq F_\alpha(s-1, rs(t-1))\}$$

在给定显著性水平α下，因素A,B的交互作用对试验结果有显著影响的拒绝域为

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

将上述结果总结，可以得到如下表内容:

表5.13双因素等重复试验的方差分析表

方差来源	平方和	自由度	均　方	F　比
因素　A	Q_A	$r-1$	$\bar{Q}_A=\dfrac{Q_A}{r-1}$	$F_A=\dfrac{\bar{Q}_A}{\bar{Q}_E}$
因素　B	Q_B	$s-1$	$\bar{Q}_B=\dfrac{Q_B}{s-1}$	$F_B=\dfrac{\bar{Q}_B}{\bar{Q}_E}$
交互作用	$Q_{A\times B}$	$(r-1)(s-1)$	$\bar{Q}_{A\times B}=\dfrac{Q_{A\times B}}{(r-1)(s-1)}$	$F_{A\times B}=\dfrac{\bar{Q}_{A\times B}}{\bar{Q}_E}$
误　　差	Q_E	$rs(t-1)$	$\bar{Q}_E=\dfrac{Q_E}{rs(t-1)}$	
总　　和	Q_T	$rst-1$		

为了计算方便，通常可以采用如下公式：令

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

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

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

则

$$\begin{cases} Q_A=U-P,\ \ Q_B=V-P,\ \ Q_{A\times B}=R-U-V-P, \\ Q_E=W-R,\ \ Q_T=W-P \end{cases}$$

例2(p171例5.6) 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各取各取4种水平，每组组合水平重复试验两次，得到下表数据问收缩率和总拉伸倍数以及这两者的交互作用对纤维弹性是否有显著的影响, 取显著性水平为0.05.

因素A / 因素B	B_1	B_2	B_3	B_4
A_1	71,73	72,73	75,73	77,75
A_2	73,75	76,74	78,77	74,74
A_3	76,73	79,77	74,75	74,73
A_4	75,73	73,72	70,71	69,69

解 $r=4, s=4, t=2, rst=32,$经计算

$$Q_T = 180.219,\quad Q_A = 70.594,\quad Q_B = 8.594,\quad Q_{A\times B} = 79.531$$

$$Q_E = 21.500$$

$$F_A = \frac{\bar{Q}_A}{\bar{Q}_E} = 17.5 > F_{0.05}(3,16) = 3.24$$

$$F_B = \frac{\bar{Q}_B}{\bar{Q}_B} = 2.1 < F_{0.05}(3,16) = 3.24$$

$$F_{A\times B} = \frac{\bar{Q}_{A\times B}}{\bar{Q}_E} = 6.6 > F_{0.05}(9,16) = 2.54$$

因而纤维收缩率对弹性是有显著影响，总拉伸倍数对弹性无显著影响，而它们的相互作用对弹性有显著影响.

三、小结

1.双因素非重复试验的方差分析步骤

(1)建立数学模型;

(2)分解平方和;

(3)研究统计特性;

(4)确定拒绝域.

2.双因素等重复试验的方差分析步骤

(1)建立数学模型;

(2)分解平方和;

(3)研究统计特性;

(4)确定拒绝域.

Thank You!