第 四 章 方差分析

第一节 单因素方差分析

第二节* 双因素方差分析

第一节 单因素方差分析

一、方差分析的概念

二、单因素方差分析问题

三、单因素方差分析方法

四、检验方法

一、方差分析的概念

- **(Analysis of variance简称ANOVA)**
- 前面已经讨论两个均值相等的假设检验问题，但在统计分析与决策中，常常遇到两个以上均值是否相等的判断与检验问题，这便需要通过方差分析来解决.
- 方差分析是一种实用的统计方法，一般按误差来源作平方和分解，再将因素平方和、误差平方和、总平方和作比较分析.

方差分析又叫变异(差异)分析，**1928**年由英国统计学家**Ronald Fisher**爵士首先提出来的，在无效假设成立的前提下检验统计量**F**的分布规律，当时的检验统计量记做**Z**，**1934**年，**W．Ｓ．Snedecor**转换了另一个更容易计算的统计量，为了纪念**Fisher**的贡献，将检验统计量命名为*F*，所以方差分析又叫*F*检验。

方差分析的用途

- 用于两个或多个均数间的比较
- 分析两个或多个因素的交互作用
- 回归方程的假设检验
- 方差齐性检验

二、单因素方差分析问题

在方差分析中，将试验结果称为试验指标.

将影响试验结果的条件称为因素.

因素在试验中所处的不同状态称为该因素的水平，只考察一个影响条件的试验称为单因素试验，相应的方差分析称为单因素方差分析.

设单因素A有a个水平$A_1, A_2, \dots, A_a$,在水平A_i (i=1,2, … ,a)下,进行n_i次独立试验，试验指标的多次测量值如下表所示.

试验指标的观察值

试验序号 因素水平	1	2	…	n_i
A_1	x_{11}	x_{12}	…	x_{1n_1}
…	…	…	…	…
A_i	x_{i1}	x_{i2}	…	x_{in_i}
…	…	…	…	…
A_a	x_{a1}	x_{a2}	…	x_{an_a}

利用假设检验方法：检验各行对应总体的均值是否相等，来回答因素*A*是否对试验指标显著的影响.

应用方差分析的前提条件

假定每行各组数据分别是来自相互独立的正态总体.

1）正态性：各水平的观察数据能够看作是从服从正态分布的总体中随机抽取的样本。

2）同方差性：各组观察数据是从具有相同方差的相互独立的总体中抽取的.

统计模型

在水平A_i (i=1,2, … ,a)下，设对应的总体的概率分布为 $X_i \sim N(\mu_i, \sigma^2)$，第$i$行对应的随机样本为

$$X_{i1},\ X_{i2},\ \ldots\ ,\ X_{in_i}$$

线性统计模型为

$$\begin{cases} X_{ij} = \mu_i + \varepsilon_{ij}\ , i = 1, 2, \cdots, a,\ j = 1, 2, \cdots, n_i \\ \varepsilon_{ij} \sim N(0, \sigma^2), \text{随机误差}\varepsilon_{ij}\text{相互独立} \end{cases}$$

总平均值 $\mu = \dfrac{1}{n}\sum_{i=1}^{a} n_i \mu_i$，其中$n = \sum_{i=1}^{a} n_i$

第i个水平A_i的效应 $\delta_i = \mu_i - \mu$， 且有$\sum_{i=1}^{a} n_i \delta_i = 0.$

三、单因素方差分析方法

方差分析方法：利用参数检验方法来检验各个总体$N(\mu_i, \sigma^2)$中各 μ_i 的是否相等.

1. 建立假设

原 假 设 H_0：$\mu_1 = \mu_2 = \dots = \mu_r$

备选假设 H_1：$\mu_i\ (i=1,2,\dots,r)$不完全相等.

等价于检验假设：

H_0：$\delta_1 = \delta_2 = \dots = \delta_a = 0$

H_1：至少有一个i，$\delta_i \neq 0$,

如何判断因素A对试验指标有是否显著的影响呢？

试验序号 / 因素水平	1	2	…	n_i
A_1	x_{11}	x_{12}	…	x_{1n_1}
…	…	…	…	…
A_i	x_{i1}	x_{i2}	…	x_{in_i}
…	…	…	…	…
A_a	x_{a1}	x_{a2}	…	x_{an_a}

在单因素方差分析问题中，除了因素A外，还要考虑随机因素对试验指标的可能影响.

若试验数据表中各个数据波动大小差异不明显，可初步判断因素A对指标没有显著影响.

若试验数据表中各个数据波动大小差异明显，此时A对指标可能有显著影响，但需要进一步用数学方法分别描述因素A的影响大小以及随机因素影响的大小.

基本思路

① 计算数据总的波动（称为变差）；

② 计算在总波动中由因素A所贡献的部分；

③ 在总波动中由随机因素所贡献的部分；

④ 比较A所贡献的部分与随机因素所贡献的部分.

2. 总变差的分解

(1) 总变差（波动）

在水平A_i下的样本 均值 $\bar{x}_{i\cdot} = \dfrac{1}{n_i}\sum_{j=1}^{n_i} x_{ij}$

样本数据的总平均值 $\bar{x} = \dfrac{1}{n}\sum_{i=1}^{a}\sum_{j=1}^{n_i} x_{ij}$

总变差 $S_T^2 = \sum_{i=1}^{a}\sum_{j=1}^{n_i}\left(x_{ij} - \bar{x}\right)^2$

反映了因素A及随机因素对试验指标的影响大小.

(2）组间变差 此处的组是指因素*A*的不同水平，它们的试验数据均值各不相同，这种差异叫**组间变差.**

可用如下 **因素*A*效应平方和** 来描述：

$$S_A^2=\sum_{i=1}^{a}\sum_{j=1}^{n_i}\left(\bar{x}_{i\bullet}-\bar{x}\right)^2=\sum_{i=1}^{a}(n_i-1)S_i^2$$

反映了因素*A*对试验指标的影响大小.

其中:

S_i^2 为因素A的第*i*水平的样本方差.

(3) 组内变差：因素A在同一同水平下试验值也大小不一，这种差异叫组内变差.

可用如下的随机误差平方和来描述

$$S_E^2 = \sum_{i=1}^{a}\sum_{j=1}^{n_i}\left(x_{ij} - \bar{x}_{i\bullet}\right)^2$$

反映了随机误差影响的大小.

(4) 总变差的分解

$$S_T^2=\sum_{i=1}^{a}\sum_{j=1}^{n_i}\left[\left(\bar{x}_{i\bullet}-\bar{x}\right)+\left(x_{ij}-\bar{x}_{i\bullet}\right)\right]^2$$

$$=\sum_{i=1}^{a}\sum_{j=1}^{n_i}\left(\bar{x}_{i\bullet}-\bar{x}\right)^2+\sum_{i=1}^{a}\sum_{j=1}^{n_i}\left(x_{ij}-\bar{x}_{i\bullet}\right)^2$$

$$+2\sum_{i=1}^{a}\sum_{j=1}^{n_i}\left(\bar{x}_{i\bullet}-\bar{x}\right)\left(x_{ij}-\bar{x}_{i\bullet}\right)$$

$$2\sum_{i=1}^{a}\sum_{j=1}^{n_i}\left(\overline{x}_{i\bullet}-\overline{x}\right)\left(x_{ij}-\overline{x}_{i\bullet}\right)$$

$$=2\sum_{i=1}^{a}\left(\overline{x}_{i\bullet}-\overline{x}\right)\sum_{j=1}^{n_i}\left(x_{ij}-\overline{x}_{i\bullet}\right)$$

$$=2\sum_{i=1}^{a}\left(\overline{x}_{i\bullet}-\overline{x}\right)\left(\sum_{j=1}^{n_i}x_{ij}-n_i\,\overline{x}_{i\bullet}\right)=0$$

$$S_T^2=\sum_{i=1}^{a}\sum_{j=1}^{n_i}\left(\overline{x}_{i\bullet}-\overline{x}\right)^2+\sum_{i=1}^{a}\sum_{j=1}^{n_i}\left(x_{ij}-\overline{x}_{i\bullet}\right)^2$$

$$S_T^2=S_A^2+S_E^2$$

$S^2{}_T$ 称为总变差;

$S^2{}_A$ 称为因素A效应平方和;

$S^2{}_E$ 称为误差平方和.

3. 统计分析

$\boxed{\dfrac{S_T^2}{\sigma^2}\text{的概率分布}}$

若H_0真，则 $x_{ij} \sim N(\mu, \sigma^2)$

即所有样本可看作来自同一正态总体.

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

其中 $S^2 = \dfrac{1}{n-1}\sum_{i=1}^{a}\sum_{j=1}^{n_i}\left(x_{ij} - \bar{x}\right)^2$为样本方差.

$\therefore \dfrac{S_T}{\sigma^2} = \dfrac{(n-1)S^2}{\sigma^2} \sim \chi^2(n-1)$, S_T^2自由度为$n-1$

3. 统计分析

$\boxed{\dfrac{S_E^2}{\sigma^2}\text{的概率分布}}$

$$S_E^2=\sum_{i=1}^{a}\sum_{j=1}^{n_i}\left(x_{ij}-\bar{x}_{i\bullet}\right)^2=\sum_{i=1}^{a}(n_i-1)S_i^2$$

其中$S_i^2=\dfrac{1}{n_i-1}\sum_{j=1}^{n_i}\left(x_{ij}-\bar{x}_{i\bullet}\right)^2$为第$i$水平下样本方差

由 $\dfrac{(n_i-1)S_i^2}{\sigma^2}\sim\chi^2(n_i-1)$ 及卡方分布的可加性可知

$$\frac{S_E^2}{\sigma^2}=\sum_{i=1}^{a}\frac{(n_i-1)S_i^2}{\sigma^2}\sim\chi^2\left(\sum_{i=1}^{a}(n_i-1)\right)=\chi^2(n-a)$$

3. 统计分析

$\dfrac{S_A^2}{\sigma^2}$的概率分布

即 $\dfrac{S_E^2}{\sigma^2} \sim \chi^2(n-a)$, S_E的自由度为$n-a$

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

$$S_A^2 = \sum_{i=1}^{a}\sum_{j=1}^{n_i}\left(\bar{x}_{i\bullet} - \bar{x}\right)^2$$

$$= \sum_{i=1}^{a} n_i\left(\bar{x}_{i\bullet} - \bar{x}\right)^2 = \sum_{i=1}^{a} n_i \bar{x}_{i\bullet}^2 - n\bar{x}^2$$

$$E\left(S_A^2\right)=\sum_{i=1}^{a} n_i E\left(\overline{x}_{i\bullet}^{2}\right)-nE\left(\overline{x}^{2}\right)$$

$$=\sum_{i=1}^{a} n_i\left(\frac{\sigma^2}{n_i}+\mu_i^2\right)-n\left(\frac{\sigma^2}{n}+\mu^2\right)$$

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

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

3. 统计分析

$\frac{S_A^2}{\sigma^2}$的概率分布

由于 $\sum_{i=1}^{a} n_i = n$， $\sum_{i=1}^{a} n_i\delta_i = 0$

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

在H_0： $\delta_i = 0$成立的条件下：

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

可证 $\frac{S_A^2}{\sigma^2} \sim \chi^2(a-1)$， S_A^2自由度为$a-1$

四、检验方法

记 $MS_A = \dfrac{S_A^2}{a-1}$， 称为$S_A$的均方

$$MS_E = \frac{S_E^2}{n-a}, \quad \text{称为} S_E \text{的均方}$$

$$F = \frac{\frac{S_A}{\sigma^2}\Big/(a-1)}{\frac{S_E}{\sigma^2}\Big/(n-a)} = \frac{MS_A}{MS_E} \sim F(a-1, n-a)$$

对于给定的α，查出$F_\alpha(a-1,n-a)$，

由样本值计算出S^2_A, S^2_E，从而算出F值。

若$F > F_\alpha(a-1,n-a)$，则拒绝H_0，

若$F < F_\alpha(a-1,n-a)$，则接受H_0。

四、检验方法

单因素方差分析表

方差来源	平方和	自由度	均方	F 比
因素A	S^2_A	$a-1$	$MS_A = \dfrac{S_A^2}{a-1}$	$F = \dfrac{MS_A}{MS_E}$
误差E	S^2_E	$n-a$	$MS_E = \dfrac{S_E^2}{n-a}$	
总和T	S^2_T	$n-1$		

简化计算与列表

简便计算公式

$$x_{i\bullet} = \sum_{j=1}^{n_i} x_{ij}\,, \quad i = 1, 2, \cdots, a$$

$$x_{\bullet\bullet} = \sum_{i=1}^{a}\sum_{j=1}^{n_i} x_{ij} = \sum_{i=1}^{a} x_{i\bullet}$$

$$S_T^2 = \sum_{i=1}^{a}\sum_{j=1}^{n_i} x_{ij}^2 - \frac{x_{\bullet\bullet}^2}{n}\,, \quad S_A^2 = \sum_{i=1}^{a} \frac{x_{i\bullet}^2}{n_i} - \frac{x_{\bullet\bullet}^2}{n}$$

$$S_E^2 = S_T^2 - S_A^2$$

单因素方差分析问题举例

例1人造纤维的抗拉强度是否受掺入其中的棉花的百分比的影响是有疑问的。现确定棉花百分比的5个水平：15%，20%，25%，30%，35%，每个水平中测5个抗拉强度的值，列于表中，问抗拉强度是否受掺入棉花的百分比的影响？（α=0.01）

棉花的百分比(i)	抗拉强度观察值（j）					
	1	2	3	4	5	$x_{i\bullet}$
15	7	7	15	11	9	49
20	12	17	12	18	18	77
25	14	18	18	19	19	88
30	19	25	22	19	23	108
35	7	10	11	15	11	54

$$x_{\bullet\bullet} = \sum_{i=1}^{5}\sum_{j=1}^{5} x_{ij} = 376$$

解

设抗拉强度为

$$x_{ij}=\mu_i+\varepsilon_{ij}\text{ ， } i,j=1,2,3,4,5$$

原假设 H_0： $\mu_1=\mu_2=\mu_3=\mu_4=\mu_5$

备选假设 H_1： $\mu_i\neq\mu_j$ 至少有一对i, j

这里$a=5$，$n_i=5,(i=1,2,3,4,5)$，$n=25$

$$S_T^2=\sum_{i=1}^{5}\sum_{j=1}^{5}x_{ij}^2-\frac{x_{\bullet\bullet}^2}{n}=7^2+7^2+\cdots+11^2-\frac{376^2}{25}$$

$$=636.96$$

$$S_A^2=\sum_{i=1}^{5}\frac{x_{i\bullet}^2}{n_i}-\frac{x_{\bullet\bullet}^2}{n}=\frac{1}{5}(49^2+\cdots+54^2)-\frac{376^2}{25}$$

$$=475.76$$

$$S_E^2=S_T^2-S_A^2=636.96-475.76=161.20$$

S_T^2,S_A^2,S_E^2的自由度分别为$24,4,20$

$$MS_A=\tfrac{475.76}{4}=118.94$$

$$MS_E=\tfrac{161.20}{20}=8.06$$

方差分析表

方差来源	平方和	自由度	均方	F 比
因素A	**475.76**	**4**	**118.94**	$\frac{118.94}{8.06}=14.76$
误差E	**161.20**	**20**	**8.06**	
总和T	**636.96**	**24**		

对于给定的$\alpha = 0.01$，查表得

$F_{\alpha}(a-1,n-a) = F_{0.01}(4,20) = 4.43$

这里$F=14.76 > 4.43 = F_{0.01}(4,20)$，

故拒绝原假设H_0，接受H_1：$\mu_i \neq \mu_j$

说明棉花的百分比对人造纤维的抗拉强度有影响。

第二节* 双因素方差分析

(无重复试验的双因素方差分析)

一、双因素方差分析

双因素方差分析

- 在许多实际问题中，往往不能只考虑一个因素的各水平的影响，还必须同时考虑几种因素的相互影响作用。研究两种因素对试验指标的影响，称为双因素方差分析。不考虑交互作用影响采用无重复双因素方差分析，考虑交互作用影响采用有重复双因素方差分析。

无重复试验的双因素方差分析

因素A有a个水平$A_1, A_2, \ldots, A_a$,

因素B有b个水平$B_1, B_2, \ldots, B_b$,

在每一个组合水平 (A_i, B_j)下,进行1次试验，

试验指标的试验值如下表所示.

1. 数据表

$B(j)$ \ $A(i)$	B_1	B_2	$\cdots$	B_j	$\cdots$	B_b	$x_{i\bullet}$
A_1	x_{11}	x_{12}	$\cdots$	x_{1j}	$\cdots$	x_{1b}	$x_{1\bullet}$
A_2	x_{21}	x_{22}	$\cdots$	x_{2j}	$\cdots$	x_{2b}	$x_{2\bullet}$
$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$
A_i	x_{i1}	x_{i2}	$\cdots$	x_{ij}	$\cdots$	x_{ib}	$x_{i\bullet}$
$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$
A_a	x_{a1}	x_{a2}	$\cdots$	x_{aj}	$\cdots$	x_{ab}	$x_{a\bullet}$
$x_{\bullet j}$	$x_{\bullet 1}$	$x_{\bullet 2}$	$\cdots$	$x_{\bullet j}$	$\cdots$	$x_{\bullet b}$	$x_{\bullet\bullet}$

2. 统计模型

设x_{ij}~$N(\mu_{ij}, \sigma^2)$，i=1,2, ... ,a, j=1,2, ... ,b

线性统计模型为

$$\begin{cases} x_{ij} = \mu_{ij} + \varepsilon_{ij} \ , i = 1,2,\cdots,a, \ j = 1,2,\cdots,b \\ \varepsilon_{ij} \sim N(0, \sigma^2), \text{随机误差} \varepsilon_{ij} \text{相互独立} \end{cases}$$

记$\mu_{ij} = \mu + \alpha_i + \beta_j$,　　其中$\mu = \dfrac{1}{ab}\sum_{i=1}^{a}\sum_{j=1}^{b} x_{ij}$

α_i为因素A的A_i水平效应, $\sum_{i=1}^{a}\alpha_i = 0$

β_j为因素B的B_j水平效应, $\sum_{j=1}^{b}\beta_j = 0$

2. 统计模型

线性统计模型变为

$$\begin{cases} x_{ij} = \mu + \alpha_i + \beta_j + \varepsilon_{ij}, \begin{pmatrix} i = 1,2,\cdots,a, \\ j = 1,2,\cdots,b \end{pmatrix} \\ \varepsilon_{ij} \sim N(0,\sigma^2), \varepsilon_{ij}\text{相互独立} \\ \sum_{i=1}^{a} \alpha_i = 0, \sum_{j=1}^{b} \beta_j = 0 \end{cases}$$

2. 统计模型

检验的假设为：

$$\begin{cases} H_{A0}: \alpha_1 = \alpha_2 = \cdots = \alpha_a = 0 \\ H_{A1}: \alpha_i \neq 0, \text{至少有一个} i \end{cases}$$

$$\begin{cases} H_{B0}: \beta_1 = \beta_2 = \cdots = \beta_b = 0 \\ H_{B1}: \beta_j \neq 0, \text{至少有一个} j \end{cases}$$

3. 总变差的分解

1) 总变异

在水平A_i下的样本均值 $\bar{x}_{i\bullet}=\dfrac{1}{b}\sum_{j=1}^{b}x_{ij}$

在水平B_j下的样本均值 $\bar{x}_{\bullet j}=\dfrac{1}{a}\sum_{i=1}^{a}x_{ij}$

样本数据的总平均值 $\bar{x}=\dfrac{1}{ab}\sum_{i=1}^{a}\sum_{j=1}^{b}x_{ij}$

总离差平方和 $S_T=\sum_{i=1}^{a}\sum_{j=1}^{b}\left(x_{ij}-\bar{x}\right)^2$

3. 总变差的分解

2) 因素间变异 各因素间的均值各不相同，

这种差异叫因素间变异.

因素A效应平方和 $$S_A^2=\sum_{i=1}^{a}\sum_{j=1}^{b}\left(\bar{x}_{i\bullet}-\bar{x}\right)^2$$

因素B效应平方和 $$S_B^2=\sum_{i=1}^{a}\sum_{j=1}^{b}\left(\bar{x}_{\bullet j}-\bar{x}\right)^2$$

3. 总变差的分解

3) 组内变异　各个因素组内的观察值也大小不一，这种差异叫组内变异，反映了随机误差的大小.

随机误差平方和

$$S_E^2 = \sum_{i=1}^{a}\sum_{j=1}^{b}[(x_{ij} - \bar{x}) - (\bar{x}_{i\bullet} - \bar{x}) - (\bar{x}_{\bullet j} - \bar{x})]^2$$

$$= \sum_{i=1}^{a}\sum_{j=1}^{b}\left(x_{ij} - \bar{x}_{i\bullet} - \bar{x}_{\bullet j} + \bar{x}\right)^2$$

3. 总变差的分解

$$S_T^2=\sum_{i=1}^{a}\sum_{j=1}^{b}\left[\left(\overline{x}_{i\bullet}-\overline{x}\right)+\left(\overline{x}_{\bullet j}-\overline{x}\right)+\left(x_{ij}-\overline{x}_{i\bullet}-\overline{x}_{\bullet j}+\overline{x}\right)\right]^2$$

$$=\sum_{i=1}^{a}\sum_{j=1}^{b}\left(\overline{x}_{i\bullet}-\overline{x}\right)^2+\sum_{i=1}^{a}\sum_{j=1}^{b}\left(\overline{x}_{\bullet j}-\overline{x}\right)^2$$

$$+\sum_{i=1}^{a}\sum_{j=1}^{b}\left(x_{ij}-\overline{x}_{i\bullet}-\overline{x}_{\bullet j}+\overline{x}\right)^2$$

$$=b\sum_{i=1}^{a}\left(\overline{x}_{i\bullet}-\overline{x}\right)^2+a\sum_{j=1}^{b}\left(\overline{x}_{\bullet j}-\overline{x}\right)^2$$

$$+\sum_{i=1}^{a}\sum_{j=1}^{b}\left(x_{ij}-\overline{x}_{i\bullet}-\overline{x}_{\bullet j}+\overline{x}\right)^2$$

3. 总变差的分解

则 $S^2_T = S^2_A + S^2_B + S^2_E$

$$S_A^2 = b\sum_{i=1}^{a}\left(\bar{x}_{i\bullet} - \bar{x}\right)^2, \text{因素}A\text{效应平方和}$$

$$S_B^2 = a\sum_{j=1}^{b}\left(\bar{x}_{\bullet j} - \bar{x}\right)^2, \text{因素}B\text{效应平方和}$$

$$S_E^2 = \sum_{i=1}^{a}\sum_{j=1}^{b}\left(x_{ij} - \bar{x}_{i\bullet} - \bar{x}_{\bullet j} + \bar{x}\right)^2, \text{随机误差平方和}$$

4. 统计分析

S^2_T的自由度为$ab-1$,

S^2_A的自由度为$a-1$,

S^2_B的自由度为$b-1$,

S^2_E的自由度为$(ab-1)-(a-1)-(b-1)$

$$= (a-1)(b-1)$$

4. 统计分析

- 均方值

$$MS_A = \frac{S_A}{a-1},$$

$$MS_B = \frac{S_B}{b-1},$$

$$MS_E = \frac{S_E}{(a-1)(b-1)}$$

4. 统计分析

- 均方值期望

$$E(MS_A)=\sigma^2+\frac{b}{a-1}\sum_{i=1}^{a}\alpha_i^2\ ,$$

$$E(MS_B)=\sigma^2+\frac{a}{b-1}\sum_{j=1}^{b}\beta_j^2\ ,$$

$$E(MS_E)=\sigma^2$$

4. 统计分析

在H_{A0}：$\alpha_i=0$，H_{B0}：$\beta_i=0$成立的条件下：可证

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

$$\frac{S_B^2}{\sigma^2} \sim \chi^2(b-1),$$

$$\frac{S_E^2}{\sigma^2} \sim \chi^2((a-1)(b-1))$$

4. 统计分析

$$F_1=\frac{\dfrac{S_A^2}{\sigma^2}\Big/(a-1)}{\dfrac{S_E^2}{\sigma^2}\Big/(a-1)(b-1)}=\frac{MS_A}{MS_E}\sim F(a-1,(a-1)(b-1))$$

$$F_2=\frac{\dfrac{S_B^2}{\sigma^2}\Big/(b-1)}{\dfrac{S_E^2}{\sigma^2}\Big/(a-1)(b-1)}=\frac{MS_B}{MS_E}\sim F(b-1,(a-1)(b-1))$$

对于给定的α,

查出$F_\alpha(a-1,(a-1)(b-1))$，$F_\alpha(b-1,(a-1)(b-1))$，

由样本值计算出S_A, S_B, S_E，从而算出F_1, F_2值.

若$F_1>F_\alpha(a-1,(a-1)(b-1))$，则拒绝$H_{A0}$，

若$F_2>F_\alpha(b-1,(a-1)(b-1))$，则拒绝$H_{B0}$.

5. 双因素方差分析表

无重复试验的双因素方差分析表

方差来源	平方和	自由度	均方	F 比
因素A	$S^2{}_A$	$a-1$	$MS_A=\frac{S_A^2}{a-1}$	$F_1=\frac{MS_A}{MS_E}$
因素B	$S^2{}_B$	$b-1$	$MS_B=\frac{S_B^2}{b-1}$	$F_2=\frac{MS_B}{MS_E}$
误差E	$S^2{}_E$	$(a-1)\cdot(b-1)$	$MS_E=\frac{S_E^2}{(a-1)(b-1)}$	
总和T	$S^2{}_T$	$ab-1$		

简化计算与列表

记

$$x_{i\bullet}=\sum_{j=1}^{b}x_{ij}\,,\quad i=1,2,\cdots,a$$

$$x_{\bullet j}=\sum_{i=1}^{a}x_{ij}\,,\quad j=1,2,\cdots,b$$

$$x_{\bullet\bullet}=\sum_{i=1}^{a}\sum_{j=1}^{b}x_{ij}=\sum_{i=1}^{a}x_{i\bullet}=\sum_{j=1}^{b}x_{\bullet j}$$

简化计算与列表

$$S_T^2=\sum_{i=1}^{a}\sum_{j=1}^{b}x_{ij}^2-\frac{x_{\bullet\bullet}^2}{ab},$$

$$S_A^2=\sum_{i=1}^{a}\frac{x_{i\bullet}^2}{b}-\frac{x_{\bullet\bullet}^2}{ab}$$

$$S_B^2=\sum_{j=1}^{b}\frac{x_{\bullet j}^2}{a}-\frac{x_{\bullet\bullet}^2}{ab}$$

$$S_E^2=S_T^2-S_A^2-S_B^2$$

双因素方差分析问题举例

例1 使用4种燃料，3种推进器作火箭射程试验，每一种组合情况做一次试验得火箭射程(单位：海里)列于表，试分析各种燃料(A_i)与各种推进器(B_j)对火箭射程有无显著影响？（α=0.05）

B_j / A_i	B_1	B_2	B_3	$x_{i\bullet}$
A_1	582	562	653	1797
A_2	491	541	516	1548
A_3	601	709	392	1702
A_4	758	582	487	1827
$x_{\bullet j}$	2432	2394	2048	$6874=x_{\bullet\bullet}$

解

这是双因素试验，不考虑交互作用.

设火箭射程为

$x_{ij} = \mu + \alpha_i + \beta_j + \varepsilon_{ij}$， i=1,2,3,4, j=1,2,3

原假设H_{A0}： $\alpha_1 = \alpha_2 = \alpha_3 = \alpha_4 = 0$

H_{B0}： $\beta_1 = \beta_2 = \beta_3 = 0$

对立假设H_{A1}： $\alpha_I \neq 0$ 至少有一个i

H_{B1}： $\beta_j \neq 0$ 至少有一个j

这里a=4，b=3，ab=12

$$S_T^2=\sum_{i=1}^{4}\sum_{j=1}^{3}x_{ij}^2-\frac{x_{\bullet\bullet}^2}{12}=582^2+562^2+\cdots+487^2$$

$$-\frac{6874^2}{12}=111342$$

$$S_A^2=\sum_{i=1}^{4}\frac{x_{i\bullet}^2}{3}-\frac{x_{\bullet\bullet}^2}{12}=\frac{1}{3}(1797^2+1548^2+1702^2+1827^2)$$

$$-\frac{6874^2}{12}=15759$$

$$S_B^2=\sum_{j=1}^{3}\frac{x_{\bullet j}^2}{4}-\frac{x_{\bullet\bullet}^2}{12}=\frac{1}{4}(2432^2+2394^2+2048^2)$$

$$-\frac{6874^2}{12}=22385$$

$$S_E^2=S_T^2-S_A^2-S_B=111342-15759-22385=73198$$

S_T^2,S_A^2,S_B^2,S_E^2的自由度分别为$11,3,2,6$

火箭射程方差分析表为

方差来源	平方和	自由度	均方	F 比
燃料A	**15759**	**3**	**5253**	**0.43**
推进器B	**22385**	**2**	**11192.5**	**0.92**
误差E	**73198**	**6**	**12199.7**	
总和T	**111342**	**11**		

对于给定的α=0.05，查表得

$F_{0.05}(3,6)=4.76$, $F_{0.05}(2,6)=5.14$

因为$F_1=0.43<4.76$, $F_2=0.92<5.14$，

所以接受原假设H_{A0}、H_{B0}，

故不同的燃料、不同的推进器对火箭射程

均无显著影响.