Ch6 方差分析

和

正交试验设计

主要内容

- ❖ 方差分析概述
- ❖ 单因子方差分析
- ❖ 不考虑交互作用的双因子方差分析
- ❖ 考虑交互作用的双因子方差分析
- ❖ 正交试验设计

方差分析概述

ANOVA 由英国统计学家 ***R.A.Fisher*** 首创.

为纪念 ***Fisher***，方差分析又称 ***F*** 检验 .

方差分析是应用广泛的实验数据统计分析方法

其实质是检验多个变量均值的一致性.

❖ 方差分析的几个基本概念

试验指标 — 试验所考察的事项，亦称响应变量.

如考察化工生产中产品的质量、数量.

试验因子 — 影响试验指标的因素.

如原料成分、原料剂量、催化剂 等.

因子水平 — 试验中因子所处的不同状态.

单因子方差分析

— 单指标、单因子、多因子水平的实验数据分析.

双因子方差分析

— 单指标、双因子、多因子水平的实验数据分析.

常用大写英文字母表示试验因子，

用大写字母加下标表示该因子的不同水平.

❖ **方差分析的工作目标**

依据实验数据，判断不同因子水平下指标变量取值否有差异？

即：比较不同因子水平下指标的均值是否相等

例1 一位英语教师想检查 3 种不同教学方法的效果，为此随机选取 24 位学生并把他们分成 3组，相应用三种方法教学. 一段时间后教师对这 24 位学生进行统考，统考成绩如表 1.

表1 英语成绩表

方法	学习成绩								
A_1	73	66	89	82	43	80	63		
A_2	88	78	91	76	85	84	80	96	
A_3	68	79	71	71	87	68	59	76	80

试问在 0.05 显著性水平下，这三种教学方法有无显著性差异？

分析 问题是判断教师所采用的不同教学方法对学生英语成绩是否有显著影响. 若有影响，哪一种教学方法好？

指标变量 — 英语成绩.

影响因子 — 教学方法，记为 A .

因子水平 —三种不同的教学方法, 记为 A_1, A_2, A_3 .

显然，这是一个单因子方差分析问题.

❖ 方差分析的三个基本的理论假设

方差分析中将因子的各个水平下的试验指标看作是不同的变量. 如三种教学方法下的英语成绩 ξ_1, ξ_2, ξ_3 .

通常假定每一个变量服从方差相等的正态分布，并且是相互独立的.

❖ 方差分析的基本思路

判断不同的教学方法对英语成绩的影响是否有显著差异，按 ***Fisher*** 的思路可以通过分析造成成绩数据差异的原因来得到答案.

实验数据（英语成绩）差异的来源：

❶ 条件误差 —由因子的不同水平(三种不同的教学方法)引起的差异.

❷ 随机误差 —由随机因素(不可控制或不可预知的因素，如考试时的环境、时间对学生的影响)引起的差异.

实验数据之间的差异可用离差平方和的概念描述.

方差分析中的一个最基本的关系式，就是将数据总的离差平方和按照产生的原因进行分解，得到：

总离差平方和 = 条件误差平方和 + 随机误差平方和

方差分析的任务就是进一步判断：

在总离差平方和中，条件误差平方和与随机误差平方和究竟哪一个是决定性的（占更大的比重）？

如果是条件误差平方和是决定性的，则说明因子的不同水平对指标变量的影响是有差异的.

因子的不同水平对指标变量的影响是否有差异，可以用不同水平下指标变量的均值是否相同来描述.

❖ 方差分析基本任务的统计描述

判断不同水平下指标变量的均值是否相同，可归结为统计推断问题，即检验假设

H_0： 不同因子水平下指标变量的均值相同;

H_1： 不同因子水平下指标变量的均值不完全相同.

围绕 H_0 的检验，应思考并解决如下三个方面的问题：

❶ 当前的问题是否满足三个基本的理论假设;

❷ 检验统计量的构造与分布是怎样的？如何决策？

❸ 若拒绝 H_0，如何找出因子的最优处理？

单因子方差分析

数据结构式

❖ 单因子方差分析统计模型

在单因子试验中，设因子 $\boldsymbol{A}$ 有 $\boldsymbol{r}$ 个水平，第 $\boldsymbol{i}$ 个水平 $\boldsymbol{A_i}$ 下指标变量 ξ_i 的均值为 μ_i，容量为 $\boldsymbol{m_i}$ 的一个样本为:

$$X_{i1}, X_{i2}, \cdots, X_{im_i}, \ i=1,2,\cdots,r. \quad (\text{与总体}\xi_i \sim N(\mu_i, \sigma^2)\text{同分布})$$

于是，单因子方差分析的统计模型为:

$$\begin{cases} X_{ij} = \mu_i + \varepsilon_{ij}, \quad i=1,2,\cdots,r, \ j=1,2,\cdots,m_i, \\ \boldsymbol{\varepsilon_{ij} \sim N(0,\sigma^2)} \text{ 且相互独立 .} \end{cases}$$

随机误差

检验假设 $\boldsymbol{H_0}$: $\boldsymbol{\mu_1 = \mu_2 = \cdots = \mu_r}$.

❖ 检验统计量的构造与分布

记：$$\boldsymbol{n=\sum_{i=1}^{r} m_i}$$ — 样本总容量

$$\bar{X}_i = \frac{1}{m_i}\sum_{j=1}^{m_i} X_{ij}$$ —水平A_i的均值

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{r}\sum_{j=1}^{m_i} X_{ij}$$ —样本总均值

$$SS_T = \sum_{i=1}^{r}\sum_{j=1}^{m_i}(X_{ij}-\bar{X})^2$$ —总离差平方和

$$\boldsymbol{f_T = n-1}$$ — SS_T的自由度

$$SS_A=\sum_{i=1}^{r}m_i(\bar{X}_i-\bar{X})^2$$ —因子A的(离差)平方和

$$f_A=r-1$$ — SS_A 的自由度

SS_A 反映组间由效应不同引起的数据差异.

$$SS_e=\sum_{i=1}^{r}SS_i=\sum_{i=1}^{r}\sum_{j=1}^{m_i}(X_{ij}-\bar{X}_i)^2$$ —误差平方和

$$f_E=n-r$$ — SSe 的自由度

$$SS_i=\sum_{j=1}^{m_i}(X_{ij}-\bar{X}_i)^2$$ —因子水平A_i的平方和

SSe 反映组内数据的随机误差.

在单因素方差分析模型中，对上述定义的三个离差平方和 SS_T，SSe，SS_A，有下面的定理：

（1）平方和分解定理

$$SS_T = SSe + SS_A$$

（2）SSe 的分布定理

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

（3）SS_A 的分布定理 当假设 H_0 为真时，有

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

(4) 检验统计量及其分布定理 当假设 H_0 为真时，有 SS_A与 SSe 相互独立, 且:

$$F=\frac{SS_A/(r-1)}{SSe/(n-r)}\sim F(r-1,n-r)\ .$$

显然，检验假设 H_0 不真时，SS_A 会变得偏大. 但是无论 H_0 真否，$SSe/(n-r)$ 都是 σ^2 的无偏估计

因此，统计量 F 的值偏大不利于H_0.

因此，可采用统计量 F 来检验假设 H_0，拒绝域为

$$W_1=\left\{F\geq F_{1-\alpha}(r-1,n-r)\right\}\ .$$

拒绝假设 H_0 的最小显著性概率(即检验的p值)为

$$p=P\big(F(r-1,n-r)\geq F\big)\ .$$

❖ 检验结果的报告

单因子方差分析表

偏差来源	离差平方和	自由度	均方和	F值	p值
因子A	SS_A	$r-1$	$SS_A/r-1$	$\frac{SS_A/r-1}{SSe/n-r}$	$P(F \geqslant F_{EST})$
误差E	SSe	$n-r$	$SSe/n-r$		
总和T	SS_T	$n-1$			

单因子方差分析主要结论的证明

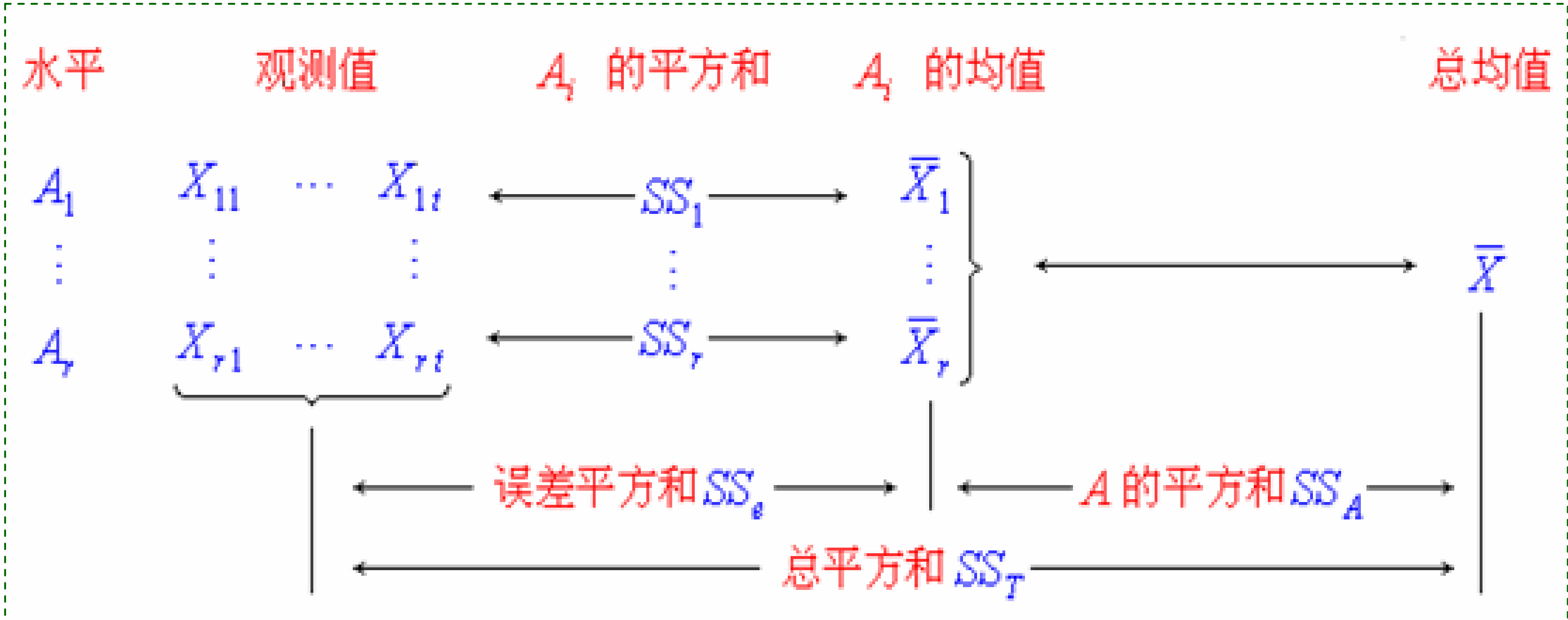
为方便起见,我们只对不同因子水平下对指标的观测次数相同的情形,来证明离差分解公式和统计量分布的结论.

不同因子水平下,样本容量不同的情形,其结论同理可证

要证明的结论:

$$\boldsymbol{SS_T = SSe + SS_A} \qquad \text{离差分解公式}$$

$$F = \frac{SS_A/(r-1)}{SSe/(n-r)} \sim F(r-1, n-r)\ . \quad (H_0\text{成立时})$$



$SS_T=\sum_{i=1}^{r}\sum_{j=1}^{t}(X_{ij}-\bar{X})^2$ 为**总平方和**，

$SS_e=\sum_{i=1}^{r}\sum_{j=1}^{t}(X_{ij}-\bar{X}_i)^2=\sum_{i=1}^{r}SS_i$ 为**误差平方和**

$SS_A=t\sum_{i=1}^{r}(\bar{X}_i-\bar{X})^2$ 为**因子 A 的平方和**。

$\bar{X}_i=\frac{1}{t}\sum_{j=1}^{t}X_{ij}$ 为**水平 A_i 的均值**，

$SS_i=\sum_{j=1}^{t}(X_{ij}-\bar{X}_i)^2$ 为**水平 A_i 的平方和**

$\bar{X}=\frac{1}{n}\sum_{i=1}^{r}\sum_{j=1}^{t}X_{ij}=\frac{1}{r}\sum_{i=1}^{r}\bar{X}_i$ 为**总均值**，

$SS_T = SSe + SS_A$ 离差分解公式的证明：

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

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

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

$$= SS_e + 2\sum_{i=1}^{r}(\sum_{j=1}^{t}X_{ij} - t\overline{X}_i)(\overline{X}_i - \overline{X}) + t\sum_{i=1}^{r}(\overline{X}_i - \overline{X})^2$$

$$= SS_e + 0 + SS_A = SS_e + SS_A$$

定理： 对单因子方差分析的模型：

设因子水平A_i下的指标值 $\xi_i \sim N(\mu_i, \sigma^2)$ 且相互独立，

对因子水平A_i下的指标作重复观测,得ξ_i的样本 $X_{i1}, X_{i2}, \ldots, X_{it}$

检验假设 H_0： $\mu_1 = \mu_2 = \cdots = \mu_r$.

则当H_0成立时， $F = \dfrac{SS_A/(r-1)}{SSe/(n-r)} \sim F(r-1, n-r)$

证： 设$\mu_1 = \mu_2 = \cdots = \mu_r = \mu$ ， 这时有$\xi_i \sim N(\mu, \sigma^2)$ ， $i = 1, 2, \cdots, r$

所以$X_{ij} \sim N(\mu, \sigma^2), \dfrac{X_{ij} - \mu}{\sigma} \sim N(0,1), i = 1, 2, \cdots, r, j = 1, 2, \cdots, t$,相互独立

$$Q=\sum_{i=1}^{r}\sum_{j=1}^{t}\left(\frac{X_{ij}-\mu}{\sigma}\right)^2=\frac{\sum_{i=1}^{r}\sum_{j=1}^{t}(X_{ij}-\bar{X}+\bar{X}-\mu)^2}{\sigma^2}$$

$$=\frac{\sum_{i=1}^{r}\sum_{j=1}^{t}(X_{ij}-\bar{X})^2}{\sigma^2}+\frac{2\sum_{i=1}^{r}\sum_{j=1}^{t}(X_{ij}-\bar{X})(\bar{X}-\mu)}{\sigma^2}+\frac{\sum_{i=1}^{r}\sum_{j=1}^{t}(\bar{X}-\mu)^2}{\sigma^2}$$

$$=\frac{SS_T}{\sigma^2}+0+\frac{n(\bar{X}-\mu)^2}{\sigma^2}$$

$$=\frac{SS_A}{\sigma^2}+\frac{SS_e}{\sigma^2}+\left(\frac{\bar{X}-\mu}{\sigma}\sqrt{n}\right)^2=Q_1+Q_2+Q_3$$

$Q_1=\dfrac{SS_A}{\sigma^2}=\dfrac{t\sum_{i=1}^{r}(\bar{X}_i-\bar{X})^2}{\sigma^2}$是$r$项的平方和，但

$\sum_{i=1}^{r}(\bar{X}_i-\bar{X})=\sum_{i=1}^{r}\bar{X}_i-r\bar{X}=0$，所以，$Q_1$的自由度$f_1=r-1$

$$Q_2=\frac{SS_e}{\sigma^2}=\frac{\sum_{i=1}^{r}\sum_{j=1}^{t}(X_{ij}-\bar{X}_i)^2}{\sigma^2}$$ 是$n=rt$项的平方和，但这n 项又满足

$\sum_{j=1}^{t}(X_{ij}-\bar{X}_i)=\sum_{j=1}^{t}X_{ij}-t\bar{X}_i=0$，$i=1,2,\cdots,r$，所以$Q_2$的自由度$f_2=n-r$

$Q_3=\left(\frac{\bar{X}-\mu}{\sigma}\sqrt{n}\right)^2$是1项的平方和，所以，$Q_3$ 的自由度$f_3=1$

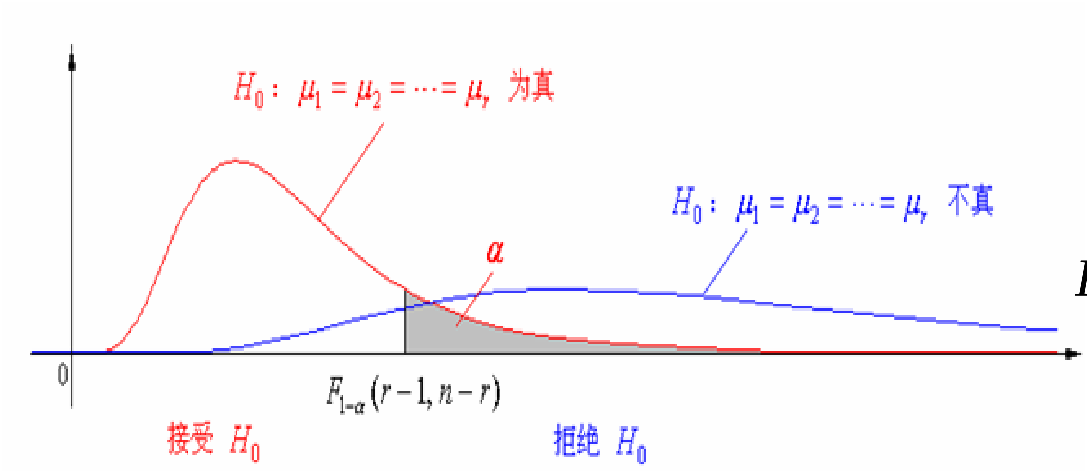
因为$f_1+f_2+f_3=(r-1)+(n-r)+1=n$ ，所以

$$Q_1=\frac{SS_A}{\sigma^2}\sim\chi^2(r-1),\quad Q_2=\frac{SS_e}{\sigma^2}\sim\chi^2(n-r),Q_3=\left(\frac{\bar{X}-\mu}{\sigma}\sqrt{n}\right)^2\sim\chi^2(1)$$

而且$Q_1 = \frac{SS_A}{\sigma^2}$， $Q_2 = \frac{SS_e}{\sigma^2}, Q_3 = \left(\frac{\bar{X}-\mu}{\sigma}\sqrt{n}\right)^2$ 相互独立

由F分布的定义可知$F_A = \frac{SS_A/(r-1)}{SS_e/(n-r)} = \frac{\frac{SS_A}{\sigma^2}\Big/(r-1)}{\frac{SS_e}{\sigma^2}\Big/(n-r)} \sim F(r-1, n-r)$

若H_0： $\mu_1 = \mu_2 = \cdots = \mu_r$不真，则$SS_A$ 的值会偏大，F_A 的值也会偏大



$F_A > F_{1-\alpha}(r-1, n-r)$时，拒绝$H_0$

单因子方差分析的计算步骤

水平	观测值	$\bar{X}_i = \frac{1}{t}\sum_{j=1}^{t} X_{ij}$	$SS_i = \sum_{j=1}^{t}(X_{ij} - \bar{X}_i)^2$
A_1 $\vdots$ A_r	$X_{11} \quad \cdots \quad X_{1t}$ $\vdots \qquad\qquad \vdots$ $X_{r1} \quad \cdots \quad X_{rt}$	$\bar{X}_1$ $\vdots$ $\bar{X}_r$	SS_1 $\vdots$ SS_r
		$SS_A = t\sum_{i=1}^{r}(\bar{X}_i - \bar{X})^2$	$SS_e = \sum_{i=1}^{r} SS_i$

来源	平方和	自由度	均方	F 值	分位数
A	SS_A	$r-1$	$MS_A = SS_A/(r-1)$	$F_A = \frac{MS_A}{MS_e}$	$F_{1-\alpha}(r-1, n-r)$
误差	SS_e	$n-r$	$MS_e = SS_e/(n-r)$		
总和	SS_T	$n-1$			

例1 为了研究不同的肥料品种对小麦产量的影响，对4种不同品种的肥料各做4次试验，得到小麦亩产量

肥料品种	亩产量			
A_1	198	196	190	166
A_2	160	169	167	150
A_3	179	164	181	170
A_4	190	170	179	188

问：肥料品种对小麦亩产量有无显著影响？（显著水平$\alpha = 0.05$）

解 这可以看作是一个单因子方差分析问题。肥料品种就是因子A，设施用4种不同肥料的小麦亩产量分别为$\xi_i \sim N(\mu_i, \sigma^2), i = 1, 2, 3, 4$。检验肥料品种对小麦亩产量有无显著影响，相当于要检验假设

$$H_0 : \mu_1 = \mu_2 = \mu_3 = \mu_4$$

计算结果见下表：

水平	观测值（亩产量）				$\bar{X}_i$	SS_i
A_1	198	196	190	166	187.50	651.00
A_2	160	169	167	150	161.50	221.00
A_3	179	164	181	170	173.50	189.00
A_4	190	170	179	188	181.75	252.75
					$SS_A = 1527.19$	$SS_e = 1313.75$

方差分析表为：

来源	平方和	自由度	均方	F 值	分位数
A	$SS_A = 1527.19$	$r-1=3$	509.06	$F_A = 4.65$	$F_{0.95}(3,12) = 3.49$
误差	$SS_e = 1313.75$	$n-r=12$	109.48		
总和	$SS_T = 2840.94$	$n-1=15$			

$F_A = 4.65 > 3.49$,所以拒绝$H_0: \mu_1 = \mu_2 = \mu_3 = \mu_4$, 结论是：肥料品种对小麦亩产量有显著影响

双因子方差分析

设 A 与 B 是对试验指标有影响的两个因子，判断因子的不同水平对试验指标的影响是否有显著差异？

双因子方差分析方法的特殊效力在于分析者可以考虑两个因子之间的交互作用对试验指标的影响.

6.2　无交互作用的双因子方差分析

设因子A有 r 个水平，因子 B 有 s 个水平，现对因子 A、B 的不同水平的每种组合下进行试验或抽样，共 $r \times s$ 个试验结果.

假定两个因子之间的不存在交互作用.

❖ 数据结构

在因子间无交互作用的情形下，可对因子水平的每个组合 (A_i, B_j) 仅安排一次试验来获取试验指标 ξ_{ij} 的测量值.

无交互作用的双因子方差分析的数据结构如下：

		因 子 B			
		B_1	B_2	…	B_s
因子 A	A_1	X_{11}	X_{12}	…	X_{1s}
	A_2	X_{21}	X_{22}	…	X_{2s}
	…	…	…		…
	A_r	X_{r1}	X_{r2}	…	X_{rs}

❖ 统计模型与检验假设

设因子组合水平 (A_i, B_j) 下的指标值 ξ_{ij} 为独立，等方差的正态总体在每一组合水平下只做一次观测,观测结果 X_{ij} 为 ξ_{ij} 的样本，即 $X_{ij} \sim N(\mu_{ij}, \sigma^2)$ 且相互独立 ， $i\ (j) = 1, 2, \cdots, r\ (s)$.

记 $\mu = \frac{1}{rs}\sum_{i=1}^{r}\sum_{j=1}^{s}\mu_{ij}$ ------总的期望平均值

$\alpha_i = \frac{1}{s}\sum_{j=1}^{s}(\mu_{ij} - \mu)$ — A_i 的效应.

$\beta_j = \frac{1}{r}\sum_{i=1}^{r}(\mu_{ij} - \mu)$ — B_j 的效应.

$\gamma_{ij} = \mu_{ij} - \mu - \alpha_i - \beta_j$ ----(A_i, B_j)的交互作用

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

于是，对无交互作用的双因子方差分析，有：

$$\gamma_{ij} = \mu_{ij} - \mu - \alpha_i - \beta_j = 0 \quad i\ (j) = 1, 2, \cdots, r\ (s),$$

即: $\mu_{ij} = \mu + \alpha_i + \beta_j$

检验假设

检验因子A的作用是否显著,相当于检验:

$$H_{01}:\ \alpha_1 = \alpha_2 = \cdots = \alpha_r$$

检验因子B的作用是否显著,相当于检验:

$$H_{02}:\ \beta_1 = \beta_2 = \cdots = \beta_s$$

❖ 检验统计量的构造与分布

检验统计量的构造思想方法类似于单因子方差分析.

记 $\bar{X}_{i\bullet} = \dfrac{1}{s}\sum_{j=1}^{s} X_{ij}$ —因子水平 A_i 的样本均值

$\bar{X}_{\bullet j} = \dfrac{1}{r}\sum_{i=1}^{r} X_{ij}$ —因子水平 B_j 的样本均值

$\bar{X} = \dfrac{1}{rs}\sum_{i=1}^{r}\sum_{j=1}^{s} X_{ij}$ —样本总均值

$SS_T = \sum_{i=1}^{r}\sum_{j=1}^{s}(X_{ij} - \bar{X})^2$ —总偏差平方和

$f_T = rs - 1$ — SS_T 的自由度

$$SS_A = s\sum_{i=1}^{r}(\bar{X}_{i\cdot} - \bar{X})^2$$ —因子 **A** 的(偏差)平方和

$$f_A = r - 1$$ — SS_A 的自由度

$$SS_B = r\sum_{i=1}^{s}(\bar{X}_{\cdot j} - \bar{X})^2$$ —因子 **B** 的(偏差)平方和

$$f_B = s - 1$$ — SS_B 的自由度

SS_A和 SS_B 反映组间由效应不同引起的数据差异.

$$SS_e = \sum_{i=1}^{r}\sum_{j=1}^{s}(X_{ij} - \bar{X}_{i\cdot} - \bar{X}_{\cdot j} + \bar{X})^2$$ —误差平方和

$$f_E = rs - r - s + 1$$ — *SSe* 的自由度

SSe 反映组内数据的随机误差.

在无交互作用的双因子方差分析模型中，有如下定理：

(1) 平方和分解定理

$$SS_T = SS_A + SS_B + SS_e \qquad\qquad f_T = f_A + f_B + f_E .$$

(2) 平方和分布定理

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

$$\frac{SS_A}{\sigma^2} \overset{H_{01}\text{真}}{\sim} \chi^2(r-1)$$

$$\frac{SS_B}{\sigma^2} \overset{H_{02}\text{真}}{\sim} \chi^2(s-1)$$

SS_A，SS_B，SS_e 相互独立

(3) 检验统计量的分布定理

$$F_A = \frac{SS_A / f_A}{SS_e / f_E} \overset{H_{01}\text{真}}{\sim} F(f_A, f_E)$$

$$F_B = \frac{SS_B / f_B}{SS_e / f_E} \overset{H_{02}\text{真}}{\sim} F(f_B, f_E)$$

❖ **检验的决策准则**

❶ 若 $\boldsymbol{F_A > F_\alpha(f_A, f_E)}$，则拒绝 $\boldsymbol{H_{01}}$.

❷ 若 $\boldsymbol{F_B > F_\alpha(f_B, f_E)}$，则拒绝 $\boldsymbol{H_{02}}$.

❖ 检验结果的报告

统计软件通常用无交互作用的双因子方差分析表报告检验结果.

无交互作用的双因子方差分析表

方差来源	平方和	自由度	均方差	F 值	p 值
因子 *A*	SS_A	$f_A = r-1$	$\frac{SS_A}{f_A}$	$F_A = \frac{SS_A}{f_A} \Big/ \frac{SSe}{f_E}$	$p_A = P(F_A > F_{AEST})$
因子 *B*	SS_B	$f_B = s-1$	$\frac{SS_B}{f_B}$	$F_B = \frac{SS_B}{f_B} \Big/ \frac{SSe}{f_E}$	$P_B = P(F_B > F_{BEST})$
误差 *E*	SSe	$f_E = (r-1)(s-1)$	$\frac{SSe}{f_E}$		
总和 *T*	SS_T	$rs-1$			

例1 对某地平均每日精神病发病人数统计如下：

		月亮		
		B_1 月圆前	B_2 月圆时	B_3 月圆后
季节	A_1 春	15.0	18.0	14.1
	A_2 夏	11.3	13.0	11.0
	A_3 秋	7.4	13.0	8.3
	A_4 冬	10.0	9.3	12.5

问：（1）季节对精神病是否有显著的影响？（显著水平$\alpha = 0.05$）

（2）月亮对精神病是否有显著的影响？（显著水平$\alpha = 0.05$）

解 这可以看作是一个不考虑交互作用的双因子方差分析问题。季节就是因子A，春、夏、秋、冬就是水平A_1, A_2, A_3, A_4。月亮就是因子B，月圆前、月圆时、月圆后就是水平B_1, B_2, B_3。设不同时期平均每日发病人数为$\xi_{ij} \sim N(\mu_{ij}, \sigma^2)$，其中，$\mu_{ij} = \mu + \alpha_i + \beta_j$，（$i = 1, 2, 3, 4$，$j = 1, 2, 3$）。

检验季节对精神病是否有显著的影响，相当于要检验 H_{01}：$\alpha_1 = \alpha_2 = \alpha_3 = \alpha_4$。

检验月亮对精神病是否有显著的影响，相当于要检验 H_{02}：$\beta_1 = \beta_2 = \beta_3$。

计算结果见下表：

	B_1 月圆前	B_2 月圆时	B_3 月圆后	$\bar{X}_{i\bullet}$	$SS_{i\bullet}$
A_1 春	15.0	18.0	14.1	15.7000	8.3400
A_2 夏	11.3	13.0	11.0	11.7667	2.3267
A_3 秋	7.4	13.0	8.3	9.5667	18.0867
A_4 冬	10.0	9.3	12.5	10.6000	5.6600
$\bar{X}_{\bullet j}$	10.9250	13.3250	11.4750	$\bar{X}=11.9083$	$SS_B=12.6467$
$SS_{\bullet j}$	30.0275	38.2675	18.2475	$SS_A=64.7758$	$SS_e=21.7667$

方差分析表为：

来源	平方和	自由度	均方	F 值	分位数
A	64.7758	$r-1=3$	21.592	5.952	$F_{0.95}(3,6)=4.76$
B	12.6467	$s-1=2$	6.3233	1.743	$F_{0.95}(2,6)=5.14$
误差	21.7667	$(r-1)(s-1)=6$	3.6278		
总和	99.1892	$rs-1=11$			

因为 $F_A=5.952>4.76=F_{1-\alpha}(r-1,(r-1)(s-1))$，所以拒绝 H_{01}：$\alpha_1=\alpha_2=\alpha_3=\alpha_4$，季节对精神病有显著的影响；

因为 $F_B=1.743<5.14=F_{1-\alpha}(s-1,(r-1)(s-1))$，所以接受 H_{02}：$\beta_1=\beta_2=\beta_3$，月亮对精神病没有显著的影响。

6.3 有交互作用的双因子方差分析

所谓交互作用，就是因素之间的联合搭配作用对实验结果产生了影响. 一般的在有重复实验的情况下,双因子方差分析的模型为:

水平组合(A_i, B_j)下的指标$\xi_{ij} \sim N(\mu_{ij}, \sigma^2)$，$i = 1, 2, \cdots, r$，$j = 1, 2, \cdots, s$

		因子 B		
		B_1	……	B_s
因子 A	A_1	$X_{111}, \cdots, X_{11t}$	……	$X_{1s1}, \cdots, X_{1st}$
	$\vdots$	$\vdots$		$\vdots$
	A_r	$X_{r11}, \cdots, X_{r1t}$	……	$X_{rs1}, \cdots, X_{rst}$

表中X_{ijk}表示因子 A、B 在第 i、j 个水平状态下第 k 个样本观测值

例如，某种农作物有A_1，A_2 两个品种，分别施以B_1，B_2 两种不同的肥料，每种组合进行1 次试验，共进行4 次试验，得到单位面积产量如下：

		因子B（肥料）	
		B_1	B_2
因子A（品种）	A_1	64	98
	A_2	85	77

如果因子A和B的作用，即品种和肥料的作用，是简单的叠加关系

那么，从组合(A_1, B_1)改为(A_2, B_2)，产量大约可以提高到

从A_1改为A_2，产量大约可以提高85-64=21

从B_1改为B_2，产量大约可以提高98-64=34

$$64+(85-64)+(98-64)=64+21+34=119$$

在给定的显著性水平α下，检验因子A，因子B，及A与B的交互作用是否显著，相当于检验如下统计假设：

$$\begin{cases} H_{01}: & \alpha_1 = \alpha_2 = \cdots = \alpha_r \\ H_{02}: & \beta_1 = \beta_2 = \cdots = \beta_s \\ H_{03}: & \gamma_{11} = \gamma_{12} = \ldots = \gamma_{rs} \end{cases}$$

其中:

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

$$\gamma_{ij} = \mu_{ij} - \mu - \alpha_i - \beta_j \qquad \text{----}(A_i, B_j)\text{的交互作用}$$

定义：

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

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

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

$$SS_A = st\sum_{i=1}^{r} (\bar{X}_{i\cdot} - \bar{X})^2 \qquad SS_B = rt\sum_{j=1}^{s} (\bar{X}_{\cdot j} - \bar{X})^2$$

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

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

$$\boxed{SS_T = SS_A + SS_B + SS_e + SS_{AB}}$$

以上定义的符号较复杂，可用下图表示

$$
\begin{tikzcd}
 & B_1 & \cdots & B_s & A_i\text{ 的平方和} & A_i\text{ 的均值} \\
A_1 & \overline{X}_{11} & \cdots & \overline{X}_{1s} & SS_{1\bullet} \arrow[l] \arrow[r] & \overline{X}_{1\bullet} \\
\vdots & \vdots & & \vdots & \vdots & \vdots \\
A_r & \overline{X}_{r1} & \cdots & \overline{X}_{rs} & SS_{r\bullet} \arrow[l] \arrow[r] & \overline{X}_{r\bullet} \\
B_j\text{ 的平方和} & SS_{\bullet 1} \arrow[u] \arrow[d] & \cdots & SS_{\bullet s} \arrow[u] \arrow[d] & SS_T \arrow[ul] \arrow[dr] & SS_A \arrow[u] \arrow[d] \\
B_j\text{ 的均值} & \overline{X}_{\bullet 1} & \cdots & \overline{X}_{\bullet s} & SS_B \arrow[l] \arrow[r] & \overline{X}
\end{tikzcd}
$$

可以证明：

$$\boxed{F_A = \frac{SS_A/(r-1)}{SS_e/rs(t-1)} \overset{H_{01}\text{真}}{\sim} F(r-1, rs(t-1))}$$

$$\boxed{F_B = \frac{SS_B/(s-1)}{SS_e/rs(t-1)} \overset{H_{02}\text{真}}{\sim} F(s-1, rs(t-1))}$$

$$\boxed{F_{AB} = \frac{SS_{AB}/(r-1)(s-1)}{SS_e/rs(t-1)} \overset{H_{03}\text{真}}{\sim} F((r-1)(s-1), rs(t-1))}$$

有交互作用的双因子方差分析表

来源	平方和	自由度	均方	F 值	分位数
A	SS_A	$r-1$	MS_A	$F_A=\dfrac{MS_A}{MS_e}$	$F_{1-\alpha}(r-1, r\,s\,(t-1))$
B	SS_B	$s-1$	MS_B	$F_B=\dfrac{MS_B}{MS_e}$	$F_{1-\alpha}(s-1, r\,s\,(t-1))$
$A\times B$	SS_{AB}	$(r-1)(s-1)$	MS_{AB}	$F_{AB}=\dfrac{MS_{AB}}{MS_e}$	$F_{1-\alpha}((r-1)(s-1), rs(t-1))$
误差	SS_e	$r\,s(t-1)$	MS_e		
总和	SS_T	$rst-1$			

检验准则是:

当计算出 F 的值大于给定 α 的临界值时，则拒绝相应的原假设.

或者由检验的 p 值作出决策，当 $p<\alpha$ 时拒绝相应的原假设.

当$F_A > F_{1-\alpha}(r-1, r\,s\,(t-1))$时，拒绝$H_{01}$,
即认为因子A 的作用显著

当$F_B > F_{1-\alpha}(s-1, r\,s\,(t-1))$时，拒绝$H_{02}$,
即认为因子B 的作用显著

当$F_{AB} > F_{1-\alpha}((r-1)(s-1), r\,s\,(t-1))$时，拒绝$H_{03}$,
即认为交互作用$A\times B$ 显著

例 1 对某种火箭，采用 A_1，A_2 两种不同的推进器，B_1, B_2, B_3 三种不同的燃料，每种组合重复进行 2 次射程试验，试验得到的火箭射程（单位：km）数据如下：

		燃料		
		B_1	B_2	B_3
推进器	A_1	58.2,　64.2	60.1,　58.3	75.8,　71.6
	A_2	56.2,　50.4	70.9,　73.3	58.2,　51.0

问：（1）推进器的差异对射程是否有显著的影响？（显著水平 $\alpha = 0.05$）

（2）燃料的差异对射程是否有显著的影响？（显著水平 $\alpha = 0.05$）

（3）推进器与燃料的交互作用对射程是否有显著的影响？（显著水平 $\alpha = 0.05$）

解 设各种推进器与燃料组合下的火箭射程为 $\xi_{ij} \sim N(\mu_{ij}, \sigma^2)$，其中，$\mu_{ij} = \mu + \alpha_i + \beta_j + \gamma_{ij}$，

$i = 1, 2$，$j = 1, 2, 3$。

检验推进器的差异对射程是否有显著的影响，相当于要检验 H_{01}：$\alpha_1 = \alpha_2$。

检验燃料的差异对射程是否有显著的影响，相当于要检验 H_{02}：$\beta_1 = \beta_2 = \beta_3$。

检验交互作用对射程是否有显著的影响，相当于要检验 H_{03}：$\gamma_{11} = \gamma_{12} = \gamma_{13} = \gamma_{21} = \gamma_{22} = \gamma_{23}$。

方差分析表为:

来源	平方和	自由度	均方	F 值	分位数
A	66.27	$r-1=1$	66.27	5.369	$F_{0.95}(1,6)=5.99$
B	160.56	$s-1=2$	80.28	6.504	$F_{0.95}(2,6)=5.14$
$A\times B$	527.36	$(r-1)(s-1)=2$	263.68	21.362	$F_{0.95}(2,6)=5.14$
误差	74.06	$r\,s\,(t-1)=6$	12.343		
总和	828.25	$r\,s\,t-1=11$			

因为 $F_A=5.369<5.99=F_{1-\alpha}(r-1,\,r\,s\,(t-1))$，所以接受 H_{01} ，推进器的差异对射程没有显著的影响。

因为 $F_B=6.504>5.14=F_{1-\alpha}(s-1,\,r\,s\,(t-1))$ ，所以拒绝 H_{02} ，燃料的差异对射程有显著的影响。

因为 $F_{AB}=21.362>5.14=F_{1-\alpha}((r-1)(s-1),\,r\,s\,(t-1))$ ，所以拒绝 H_{03} ，推进器与燃料的交互作用对射程有显著的影响。事实上，从试验数据可以看出，要使射程尽量远，A_1 推进器最好选用 B_3 燃料，而 A_2 推进器则最好选用 B_2 燃料。