第三章 方差分析

方差分析是试验研究中分析试验数据的重要方法，应用十分广泛。

第一节 概述

一、 方差分析的必要性

在第二章中已经讨论了两个样本的正态总体均值是否相等的假设检验问题。但在生产实践中会经常遇到检验多个样本的正态总体均值是否相等的问题。

例 3-1 以淀粉为原料生产葡萄糖的过程中，残留有许多糖蜜，可作为生产酱色的原料。在生产酱色之前应进行除杂，在试验中选用了五种不同的除杂方法，每种方法重复四次，结果见表 3-1。

表 3-1 不同除杂方法的除杂量(g/Kg)

除杂方法(A_i)	除杂量(x_{ij})				平均($\overline{x}_i$)
A_1	25.6	22.2	28.0	29.8	26.4
A_2	24.4	30.0	29.0	27.5	27.7
A_3	25.0	27.7	23.0	32.2	27.0
A_4	28.8	28.0	31.5	25.9	28.6
A_5	20.6	21.2	22.0	21.2	21.3

从表 3-1 可见，各次试验结果是参差不齐的。可以认为，同一除杂方法重复试验得到的四个数据的差异是由随机误差造成的，而随机误差常常是服从正态分布的，这种除杂量应该有一个理论上的均值。而对于不同的除杂方法，除杂量应有不同的均值，这种均值之间的差异是由于除杂方法的不同而造成的。于是我们可以认为这五种除杂方法所得到的数据是来自均值不同的五个正态总体，且由于试验中其他条件相对稳定，因而可以认为每个总体的方差是相同的。这样判断除杂方法对除杂效果是否有显著影响的问题，就转化为检验五个具有相

同方差的正态总体的均值是否相同的问题，即检验假设

H_0：$\mu_1=\mu_2=\mu_3=\mu_4=\mu_5$

在上述这种情况下，第二章介绍的方法不再适用。这是因为

①、需要检验 H_0：$\mu_1=\mu_2$，$\mu_1=\mu_3$，…，$\mu_4=\mu_5$，共需检验 $k(k-1)/2=5(5-1)/2=10$ 个假设，非常繁琐。

②、样本进行两两比较时，只能由 $2(n-1)=2\times(4-1)=6$ 个自由度估计样本平均值的标准误差，而不能由 $5\times(4-1)=15$ 个自由度一起估计，精度不够高。

③. 两两检验会随样本个数的增加而大大增加 α 错误的机会，比如在两两比较中取 $\alpha=0.05$，10 次比较的结论都正确的概率为 $0.95^{10}=0.599$，至少作出一次错误结论的概率为 $1-0.95^{10}=0.401$，这时的检验结果已很不可靠。

对于这种多个总体样本均值的假设检验，需采用方差分析方法。

二、 方差分析的基本思想

方差分析的实质就是检验多个正态总体均值是否相等。从表 3-1 可见，表中的 20 个数据是参差不齐的，数据波动的可能原因来自两个方面：一是由于因素的水平不同，即除杂方法的不同造成的。二是来自偶然误差，这是由于试验中存在的偶然因素(例如环境，原材料成分，测试技术等微小而随机的变化)引起的。

我们把由试验因素的水平变化引起的试验数据波动称为条件差异；把由随机因素引起的试验数据波动称为随机误差或试验误差。方差分析就是将试验数据的总波动分解为两部分，一部分反映由条件差

异引起的波动，另一部分反映由试验误差引起的波动。然后借助 F 检验法，检验假设 H_0： $\mu_1=\mu_2=\mu_3=\mu_4=\mu_5$，从而确定因素对试验结果的影响是否显著。

第二节 单因素试验的方差分析

一、 单因素方差分析问题的一般提法

设试验所考察的因素有 m 个水平：A_1，A_2，…，A_i，…，A_m，在每个水平上重复进行 r 次试验，水平 A_i 的第 j 次试验值为 x_{ij}(i=1，2，…，m；j=1，2，…，r)，可得到试验数据和计算表的模式如表 3-2所示。要求根据试验数据判断因素对试验结果是否有显著影响。

表 3-2 单因素试验数据和计算表

水平	重复试验序号						$x_{i。}$	$\overline{x}_{i\square}$
	1	2	$\square$	j	$\square$	r		
A_1	x_{11}	x_{12}	$\square$	x_{1j}	$\square$	x_{1r}		
A_2	x_{21}	x_{22}	$\square$	x_{2j}	$\square$	x_{2r}		
⋮	⋮	⋮		⋮		⋮		
A_i	x_{i1}	x_{i2}	$\square$	x_{ij}	$\square$	x_{ir}		
⋮	⋮	⋮		⋮		⋮		
A_m	x_{m1}	x_{m2}	$\square$	x_{mj}	$\square$	x_{mr}		
总和							$x..$	$\overline{x}..$

表中：$x_{i\square}=\boxed{r}_{j=1}x_{ij}$，$\overline{x}_{i\square}=\frac{1}{r}x_{i\square}$，$x_{\square\square}=\boxed{m}_{i=1}x_{i\square}$，$\overline{x}_{\square\square}=\frac{1}{m}\boxed{m}_{i=1}x_{i\square}$

二、 单因素方差分析的前提条件

单因素方差分析是建立在下述假设的基础上：

1、在每一水平上的试验结果是一个随机变量 x_{ij}(i 为第 i 个水平，j 为第 j 次试验)，且服从正态分布。x_{i1}，x_{i2}，$\square$，x_{ir} 是第 i 个水平的正态总体中抽出的一个简单随机样本，样本容量为 r。

2、所有 m 个不同水平所对应的 m 个正态总体的方差是相等的，

具有方差齐性，$x_{ij} \sim N(\mu_i, \sigma^2)$。

3、m 个总体是相互独立的，样本与样本之间也是相互独立的，要检验的假设是：

H_0：$\mu_1 = \mu_2 = \cdots = \mu_m$

若拒绝 H_0，则可以认为至少有两个水平之间的差异是显著的，因素 A 对试验结果有显著影响；反之，若接受 H_0，则可以认为因素 A 对试验结果无显著影响，试验结果在各水平之间的不同是由随机因素引起的。

三、 单因素方差分析的一般步骤

(一)偏差平方和的分解

把整个试验所得的每一个试验值 x_{ij} 对其总平均值 $\overline{x}_{\cdot\cdot}$ 的偏差进行平方并求总和，就是总的偏差平方和，用 S_T 表示，它反映了全部试验值间的总的波动情况。

$$S_T = \sum_{i=1}^{m}\sum_{j=1}^{r}(x_{ij} - \overline{x}_{\cdot\cdot})^2 \qquad (3-1)$$

将式(3-1)进行分解

$$S_T = \sum_{i=1}^{m}\sum_{j=1}^{r}(x_{ij} - \overline{x}_{\cdot\cdot})^2$$

$$= \sum_{i=1}^{m}\sum_{j=1}^{r}[(x_{ij} - \overline{x}_{i\cdot}) + (\overline{x}_{i\cdot} - \overline{x}_{\cdot\cdot})]^2$$

$$= \sum_{i=1}^{m}\sum_{j=1}^{r}(x_{ij} - \overline{x}_{i\cdot})^2 + \sum_{i=1}^{m}\sum_{j=1}^{r}(\overline{x}_{i\cdot} - \overline{x}_{\cdot\cdot})^2$$

$$+ 2\sum_{i=1}^{m}\sum_{j=1}^{r}(x_{ij} - \overline{x}_{i\cdot})(\overline{x}_{i\cdot} - \overline{x}_{\cdot\cdot})$$

其中 $\sum_{i=1}^{m}\sum_{j=1}^{r}(x_{ij}-\bar{x}_{i\cdot})(\bar{x}_{i\cdot}-\bar{x}_{\cdot\cdot})=\sum_{i=1}^{m}(\bar{x}_{i\cdot}-\bar{x}_{\cdot\cdot})\sum_{j=1}^{r}(x_{ij}-\bar{x}_{i\cdot})=0$

所以 $S_T=\sum_{i=1}^{m}\sum_{j=1}^{r}(x_{ij}-\bar{x}_{i\cdot})^2+\sum_{i=1}^{m}\sum_{j=1}^{r}(\bar{x}_{i\cdot}-\bar{x}_{\cdot\cdot})^2$ (3-2)

令 $S_A=\sum_{i=1}^{m}\sum_{j=1}^{r}(\bar{x}_{i\cdot}-\bar{x}_{\cdot\cdot})^2=r\sum_{i=1}^{m}(\bar{x}_{i\cdot}-\bar{x}_{\cdot\cdot})^2$ (3-3)

S_A 是各条件(水平)下的平均值与总平均值的偏差平方和，它反映了因素 A 的水平变化引起的波动，称为组间偏差平方和(或因素平方和)。

令 $Se=\sum_{i=1}^{m}\sum_{j=1}^{r}(x_{ij}-\bar{x}_{i\cdot})^2$ (3-4)

Se 是各条件(水平)下的试验值与该条件下的平均值的偏差的平方和，它反映了随机误差引起的波动，称为组内偏差平方和(或误差平方和)。

因此 $S_T=S_A+Se$ (3-5)

这样就将总的偏差平方和分解成组间偏差平方和与组内偏差平方和。

(二) 偏差平方和的简化计算

为计算简便，常用下列简便算法求 S_T、S_A、Se

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

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

$$=\sum_{i=1}^{m}\sum_{j=1}^{r}x_{ij}^2-2\left(\sum_{i=1}^{m}\sum_{j=1}^{r}x_{ij}\right)\bar{x}_{\cdot\cdot}+\sum_{i=1}^{m}\sum_{j=1}^{r}\bar{x}_{\cdot\cdot}^2$$

$$=\sum_{i=1}^{m}\sum_{j=1}^{r}x_{ij}^{2}-2(mr\,\bar{x}_{\cdot\cdot})\bar{x}_{\cdot\cdot}+mr\,\bar{x}_{\cdot\cdot}^{2}$$

$$=\sum_{i=1}^{m}\sum_{j=1}^{r}x_{ij}^{2}-n\left(\frac{x_{\cdot\cdot}}{n}\right)^{2}$$

$$=\sum_{i=1}^{m}\sum_{j=1}^{r}x_{ij}^{2}-\frac{1}{n}x_{\cdot\cdot}^{2} \qquad (3\text{-}6)$$

式中：$n=mr$

令 $Q_T=\sum_{i=1}^{m}\sum_{j=1}^{r}x_{ij}^{2}$，$CT=\frac{1}{n}x_{\cdot\cdot}^{2}$ (CT 称为修正项)

则 $S_T=Q_T-CT$ (3-7)

同样 $S_A=r\sum_{i=1}^{m}(\bar{x}_{i\cdot}-\bar{x}_{\cdot\cdot})^{2}$

$$=r\sum_{i=1}^{m}(\bar{x}_{i\cdot}^{2}-2\bar{x}_{i\cdot}\bar{x}_{\cdot\cdot}+\bar{x}_{\cdot\cdot}^{2})$$

$$=r\sum_{i=1}^{m}\bar{x}_{i\cdot}^{2}-2r\bar{x}_{\cdot\cdot}\sum_{i=1}^{m}\bar{x}_{i\cdot}+r\sum_{i=1}^{m}\bar{x}_{\cdot\cdot}^{2}$$

$$=r\sum_{i=1}^{m}\bar{x}_{i\cdot}^{2}-2r\bar{x}_{\cdot\cdot}\sum_{i=1}^{m}\bar{x}_{i\cdot}+mr\,\bar{x}_{\cdot\cdot}^{2}$$

$$=\frac{1}{r}\sum_{i=1}^{m}x_{i\cdot}^{2}-\frac{1}{n}x_{\cdot\cdot}^{2}$$

令 $Q_A=\frac{1}{r}\sum_{i=1}^{m}x_{i\cdot}^{2}$，则 $S_A=Q_A-CT$ (3-8)

求出 S_T 和 S_A 后，再利用 $Se=S_T-S_A$ (3-9)

求出 Se。

(三) 计算自由度和方差(平均偏差平方和)

偏差平方和的大小，与参与求和的项数有关，为了比较 S_A 和 Se，应消除求和项数的影响，比较它们的平均值。从数学上的理论推导可

知，S_A 与 Se 的平均值，不是把 S_A 和 Se 分别除以相应的参与求和的项数，而应当除以它们的自由度。

$S_T=\sum_{i=1}^{m}\sum_{j=1}^{r}(x_{ij}-\bar{x}_{\cdot\cdot})^2$ 中有 mr＝n 个数据 x_{ij}，存在一个线性约束 $\sum_{i=1}^{m}\sum_{j=1}^{r}(x_{ij}-\bar{x}_{\cdot\cdot})=0$，因此 S_T 的自由度 $f_T=mr-1=n-1$；

$S_A=r\sum_{i=1}^{m}(\bar{x}_{i\cdot}-\bar{x}_{\cdot\cdot})^2$ 中有 m 个数据 $\bar{x}_{i\cdot}$，存在一个线性约束 $\sum_{i=1}^{m}(\bar{x}_{i\cdot}-\bar{x}_{\cdot\cdot})=0$，因此 S_A 的自由度 $f_A=m-1$；

$Se=\sum_{i=1}^{m}\sum_{j=1}^{r}(x_{ij}-\bar{x}_{i\cdot})^2$ 中有 mr＝n 个数据 x_{ij}，存在 m 个线性约束 $\sum_{j=1}^{r}(x_{ij}-\bar{x}_{i\cdot})=0$ (i＝1，2，…，m)，所以 Se 的自由度 $fe=mr-m=n-m$

显然 $f_T=f_A+fe$ (3-10)

式(3-10称为偏差平方和的自由度分解公式。总自由度 f_T 是总的数据个数减 1，组间自由度 f_A 是因素的水平数减 1。

一般先求出 f_T 和 f_A，再利用

$fe=f_T-f_A$ (3-11)

求出组内自由度 fe。下面即可求出 S_A 和 Se 的平均值：

$V_A=S_A/f_A$ (3-12)

$Ve=Se/fe$ (3-12)

V_A 和 Ve 分别称为组间方差(或均方)和组内方差(或均方)。

(四) 显著性检验

H_0：$\mu_1 = \mu_2 = \cdots = \mu_m$，若 H_0 为真，则全体样本可看作是来自同一正态总体 $N(\mu, \sigma^2)$。因此 $S_T/(n-1)$，$S_A/(m-1)$，$Se/(n-m)$ 都是总体方差 σ^2 的无偏估计值，所以比值

$$F = \frac{V_A}{Ve} = \frac{S_A/(m-1)}{Se/(n-m)} \qquad (3\text{-}14)$$

应接近于 1。如果 F 值比 1 大得多，即 V_A 明显地大于 Ve，就有理由认为原假设 H_0 不成立。即因素 A 对试验结果的影响显著。这种比较方差(或均方)大小来判断原假设 H_0 是否成立的方法，就是方差分析。

事实上，当原假设 H_0 成立时，S_A，Se 分别是自由度为 $m-1$，$n-m$ 的 χ^2 变量，因而统计量 F 服从自由度 $f_1=m-1$，$f_2=n-m$ 的 F 一分布。对于给定的显著性水平 α，和自由度 f_1、f_2，可查表得出临界值 F_α。将由样本值算出的 F 的值 F_A 和 F_α 比较，若 $F_A > F_\alpha$，则否定原假设 H_0；若 $F_A < F_\alpha$，则接受原假设 H_0，即可认为因素 A 对试验结果无显著影响。

对于显著性水平 α 的选取，应根据具体情况而定，通常取 $\alpha = 0.01$ 和 $\alpha = 0.05$，从 F 分布表上查出 $F_{0.01,(f_1, f_2)}$ 和 $F_{0.05,(f_1, f_2)}$。若 $F_A > F_{0.01,(f_1, f_2)}$，则可判定因素 A 为高度显著，记为“**”；若 $F_{0.05,(f_1, f_2)} < F_A \leqslant F_{0.01,(f_1, f_2)}$，则可判定因素 A 为显著，记为“*”；若 $F_A \leqslant F_{0.05,(f_1, f_2)}$，则判定因素 A 为不显著。

(五) 列出方差分析表

方差分析的步骤基本上就是假设检验的步骤，特殊的只是检验用的统计量是由两个方差(或均方)之比构成。具体进行方差分析时，主

要是计算这两个方差(或均方)，由于计算过程较繁，一般是把计算结果列成方差分析表，其格式如表 3-3 所示。

表 3-3 方差分析表

方差来源	偏差平方和	自由度	方差(或均方)	F 值	Fα	显著性
因素 A	S_A	m－1	V_A	$F_A = V_A/Ve$	查表	
误差 e	Se	n－m	Ve			
总和	S_T	n－1				

四、 单因素方差分析实例

现在仍以例 3-1 的试验数据为例，说明单因素方差分析的步骤。

表 3-4 试验数据计算表

除杂工艺(A_i)	试验次序 1	2	3	4	$x_{i.}$	$\sum_{j=1}^{4} x_{ij}^2$
A_1	25.6	22.2	28.0	29.8	105.6	2820.24
A_2	24.4	30.0	29.0	27.5	110.9	3092.61
A_3	25.0	27.7	23.0	32.2	107.9	2958.13
A_4	28.8	28.0	31.5	25.9	114.2	3276.13
A_5	20.6	21.2	22.0	21.2	85.0	1807.24
总和					523.6	13954.72

(一) 计算偏差平方和及自由度

$$x_{..}=523.6$$

$$CT=\frac{1}{n}x_{..}^2=\frac{1}{20}\times 523.6^2=13707.85$$

$$Q_T=\sum_{i=1}^{5}\sum_{j=1}^{4}x_{ij}^2=(2820.24+3092.61+\cdots+1807.24)=13954.72$$

$$S_T=Q_T-CT=13954.72-13707.85=246.87$$

$$f_T=n-1=20-1=19$$

$$Q_A=\frac{1}{4}\sum_{i=1}^{5}x_{i.}^2=\frac{1}{4}\times(105.6^2+110.9^2+\ldots+85.0^2)=13839.81$$

S_A　$Q_A - CT = 13839.81 - 13707.85 = 131.96$

$f_A = m - 1 = 5 - 1 = 4$

$Se = S_T - S_A = 246.87 - 131.96 = 114.91$

$fe = f_T - f_A = 19 - 4 = 15$

(二) 计算方差(或均方)

$V_A = S_A / f_A = 131.96/4 = 32.99$

$Ve = Se/fe = 114.91/15 = 7.66$

(三) 列出方差分析表

表 3-5　方差分析表

方差来源	偏差平方和	自由度	方差(或均方)	F 值	F	显著性
因素 A	$S_A = 131.96$	$f_A = 4$	$V_A = 32.99$	$F_A = 4.31$	$F_{0.05(4,15)} = 3.06$	*
误差 e	$Se = 114.91$	$fe = 15$	$Ve = 7.66$		$F_{0.01(4,15)} = 4.89$	
总和	$S_T = 246.87$	$f_T = 19$				

由于 $F_A > F_{0.05(4,15)}$，故拒绝 H_0，即不同除杂方法对除杂效果有显著影响。

在有些情况下,可用下式对原数据作线性变换,从而使计算简化：

$$x_{ij}' = \frac{x_{ij} - a}{b} \qquad (b \quad 0) \qquad (3\text{-}14)$$

用 x_{ij}' 进行方差分析，算得的 F_A 是不变的。经过线性变换后，计算方便得多，方差分析的结果不变。

第三节 双因素试验方差分析

在科研和生产实践中,常常需要同时研究两个以上因素对试验结

例如，研究不同浸提温度和浸提时间对茶叶有效成分提取率的影响。这就要对两个试验因素进行方差分析。对双因素试验的方差分析，基本思想和方法与单因素试验的方差分析相似，所不同的是在双因素试验中，有可能出现交互作用。

一、 双因素试验的方差分析

(一) 问题的一般提法

某项试验要同时考察因素 A 和 B 对试验结果的影响，因素 A 取 A_1,A_2， $,A_a$ 共 a 个水平，因素 B 取 $B_1,B_2,\cdots,B_b$ 共 b 个水平。A 和 B 两因素的每种水平搭配 $A_iB_j(i=1,2,\cdots,a;\ j=1,2,\cdots,b)$各进行一次独立试验，共进行 $a\times b=n$ 次试验，试验数据为 $x_{ij}(i=1,2,\cdots,a;j=1,2,\cdots,b)$。

表 双因素无重复试验数据和计算表

因素 A	因素 B						$x_{i\cdot}$	$\bar{x}_{i}$
	B_1	B_2	$\cdots$	B_j	$\cdots$	B_b		
A_1	x_{11}	x_{12}	$\cdots$	x_{1j}	$\cdots$	x_{1b}	$x_{1\cdot}$	$\bar{x}_{1\cdot}$
A_2	x_{21}	x_{22}	$\cdots$	x_{2j}	$\cdots$	x_{2b}	$x_{2\cdot}$	$\bar{x}_{2\cdot}$
$\vdots$	$\vdots$	$\vdots$		$\vdots$		$\vdots$	$\vdots$	$\vdots$
A_i	x_{i1}	x_{i2}	$\cdots$	x_{ij}	$\cdots$	x_{ib}	$x_{i\cdot}$	$\bar{x}_{i\cdot}$
$\vdots$	$\vdots$	$\vdots$		$\vdots$		$\vdots$	$\vdots$	$\vdots$
A_a	x_{a1}	x_{a2}	$\cdots$	x_{aj}	$\cdots$	x_{ab}	$x_{a\cdot}$	$\bar{x}_{a\cdot}$
$x_{\cdot j}$	$x_{\cdot 1}$	$x_{\cdot 2}$	$\cdots$	$x_{\cdot j}$	$\cdots$	$x_{\cdot b}$	$x_{\cdot\cdot}$	$\bar{x}_{\cdot\cdot}$

表中 $x_{i\cdot}=\sum_{j=1}^{b}x_{ij}\ (i=1,2,\cdots,a)$，$\bar{x}_{i\cdot}=\frac{1}{b}x_{i\cdot}$

$x_{\cdot j}=\sum_{i=1}^{a}x_{ij}\ (j=1,2,\cdots,b)$，$\bar{x}_{\cdot j}=\frac{1}{a}x_{\cdot j}$

以上内容仅为本文档的试下载部分，为可阅读页数的一半内容。如要下载或阅读全文，请访问：https://d.book118.com/298037013131006024