**30. 方差分析Ⅰ —原理**

**（一）方差分析概述**

**一、概述**

方差分析是对总变异进行分析。看总变异是由哪些部分组成的，这些部分间的关系如何。

方差分析，是用来检验两个或两个以上均值间差别显著性（影响观察结果的因素：原因变量（列变量）的个数大于2，或分组变量（行变量）的个数大于1）。一元时常用F检验（也称一元方差分析），多元时用多元方差分析（最常用Wilks’∧检验）。

方差分析可用于：

（1）完全随机设计（单因素）、随机区组设计（双因素）、析因设计、拉丁方设计和正交设计等资料；

（2）可对两因素间交互作用差异进行显著性检验；

（3）进行方差齐性检验。

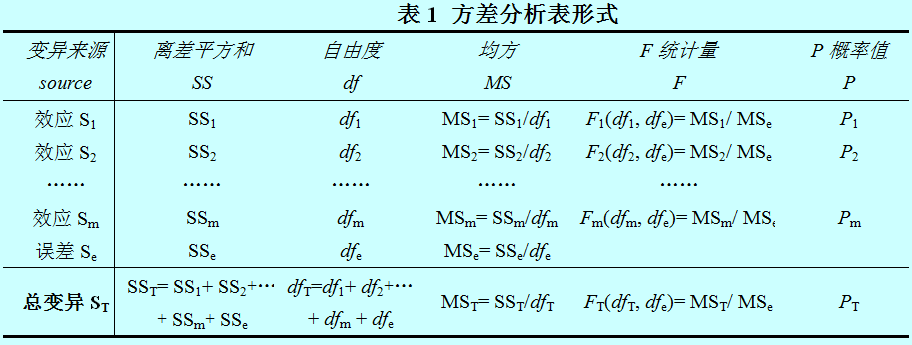
要比较几组均值时，理论上抽得的几个样本，都假定来自正态总体，且有一个相同的方差，仅仅均值可以不相同。还需假定每一个观察值都由若干部分累加而成，也即总的效果可分成若干部分，而每一部分都有一个特定的含义，称之谓效应的可加性。所谓的方差是离均差平方和除以自由度，在方差分析中常简称为均方（Mean Square）。

**二、基本思想**

基本思想是，将所有测量值上的总变异按照其变异的来源分解为多个部份，然后进行比较，评价由某种因素所引起的变异是否具有统计学意义。

根据效应的可加性，将总的**离均差平方和**分解成若干部分，每一部分都与某一种效应相对应，总自由度也被分成相应的各个部分，各部分的离均差平方除以各自的自由度得出各部分的均方，然后列出方差分析表算出F检验值，作出统计推断。

方差分析的关键是**总离均差平方和**的分解，分解越细致，各部分的含义就越明确，对各种效应的作用就越了解，统计推断就越准确。



效应项与试验设计或统计分析的目的有关，一般有：主效应（包括各种因素），交互影响项（因素间的多级交互影响），协变量（来自回归的变异项），等等。

当分析和确定了各个效应项S后，根据原始观察资料可计算出各个离均差平方和SS，再根据相应的自由度*df*，由公式MS=SS/*df*，求出均方MS，最后由相应的均方，求出各个变异项的*F*值，*F*值实际上是两个均方之比值，通常情况下，分母的均方是误差项的均方。

根据*F*值的分子、分母均方的自由度*f*1和*f*2，在确定显著性水平为α情况下，由F(*f*1, *f*2)临界值表查得单侧Fα界限值。当F<Fα时，则P值>α，不拒绝原假设H0，说明不拒绝这个效应项的效应为0的原假设，也即这个效应项是可能对总变异没有实质影响的；若F>Fα则P值≤α，拒绝原假设H0，也即这个效应项是很可能对总变异有实质影响的。

**三、方差分析的实验设计**

为了确定方差分析表中各个有关效应项，需要在试验设计阶段就作出安排，再根据设计要求进行试验，得出原始观察值，按原来设计方案算出方差分析表中的各项。

在试验设计阶段通常需要考虑如下4个方面：

（1）研究的主要变量（因变量）

即试验所要观察的主要指标，一次试验时可以有多个观察指标，方差分析时也可以同时对多个因变量进行分析；

（2）因素和水平

试验的因素（factor）可以是品种、人员、方法、时间、地区等等，因素所处的状态叫水平（level）。在每一个因素下面可以分成若干水平。

例如，某工厂的原料来自4个不同地区，那么用不同地区的原料生产的产品质量是否一致呢？所要比较的地区就是因素，4个地区便是地区这一因素的4个水平。当某个主要因素的各个水平间的主要因变量的均值呈现统计显著性时，必要时可作两两水平间的比较，称为均值间的两两比较。

（3）因素间的交互影响

多因素的试验设计，有时需要分析因素间的交互影响（interaction），2个因素间的交互影响称为一级交互影响（A×B）；3个因素间的交互影响称为二级交互影响（A×B×C）。

当交互影响项呈现统计不显著时，表明各个因素独立，当呈现统计显著时，就需要列出这个交互影响项的效应，以助于作出正确的统计推断。

**四、单因素方差分析（完全随机设计的方差分析）**

又分为两类：

**完全随机设计**——从符合条件的总体中完全随机地抽取所需数目的受试对象，再将全部受试对象完全随机地分配到k组中去。此时，受试对象与试验因素间无直接联系。

 **组内完全随机设计**——按试验因素的k个水平将全部受试对象划分成k个子总体，再分别从k个子总体中完全随机地抽取所需数目的受试对象。此时，试验因素的各水平决定了受试对象各自应该归属的组别。

设因素A有k个水平A1, …, Ak，在每一个水平下考察的指标可以看成一个总体，故有k个总体，并假定：

① 每一总体均服从正态分布；

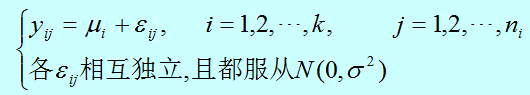
② 每一总体的方差相同；

③ 从每一总体中抽取的样本相互独立。

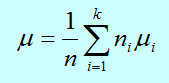
要比较各个总体的均值是否一致，就是要检验各总体的均值是否相同：

H0：μ1 = … = μk; H1：μ1 , … , μk不全相同

设从第i个总体获得容量为ni的样本观察值为yi1, …, yin\_i，各样本间相互独立。样本观察值可看成是来自均值为μi的总体，这样yij就是其均值μi与随机误差εij迭加而产生的：



为了能更仔细地描述数据，常在方差分析模型中引人一般平均与效应的概念。称各个μi的加权平均

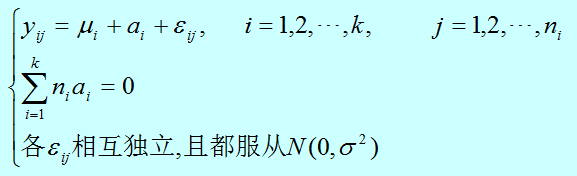


为总平均，其中. 称

ai = μi – μ, i=1,…,k

为因素A在第i水平的主效应，也简称为Ai的效应。易知.

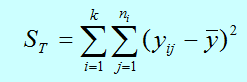
单因素方差分析的统计模型可改写为：



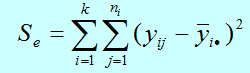
假设检验的原假设可改写为：H0：a1 = … = ak.

各yij间差异的原因可能有两个：一是假设H0不真，即各水平下总体均值μi（或水平效应ai）不同，导致从各总体中获得的样本观察值有差异；另一可能是H0为真，差异是由于随机误差εij引起的。

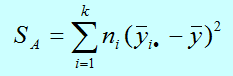
各yij间总的差异大小可用**总偏差平方和**ST表示：



随机误差引起的数据间的差异可以用组内偏差平方和（也称误差偏差平方和）Se表示：



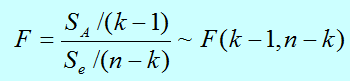
由于效应不同引起的数据差异可以用组间偏差平方和（也称因素A的偏差平方和）表示SA表示：



则总的差异=组内差异+组间差异（ST=Se+SA）. 由于



故可采用统计量



来做假设检验。

**五、多重比较**

当k组均值比较，如果经过F检验拒绝原假设，表明因素A是显著的，即k个水平对应的指标均值不全相等，但不一定两两之间都有差异。此时，还需要进一步去确认哪些水平间是确有差异的，哪些水平间无显著差异。同时比较任意两个水平均值间有无显著性差异的问题称为多重比较。即同时检验以下个假设：



多重比较的方法分为三类：临界值相对固定的两两比较、临界值不固定的多级检验、全部处理组均值与一个对照组均值比较。

每类根据所控制误差的类型和大小不同，又有许多不同的具体方法：

T（成组比较t检验法）

Bon（Bonforroni t检验法）

Dunnett（与对照组均数比较）

SNK（Student-Newman-Keuls或称q检验法）

Tukey（学生化极差HSD或称最大显著差）

Duncan（新多极差检验法）

LSD（最小显著差）

SIDAK（Sidak不等式进行校正t检验法）

SCHEFFE（Scheffe的多重对比检验）

Waller-Duncan（k比率t检验）

GT2或SMM（学生化最大模数和Sidak不等式进行校正t检验法）

REGWF（多重F检验）

REGWQ（多重极差检验）

在多重比较时，选用什么样的检验方法，首先要注意每种方法适用的试验设计条件，其次要关心所要控制的误差类型和大小。

例如，某因素有10个水平，若采用通常的t检验进行多重比较，共需要比较的次数为次，即使每次比较时都把第一类错误α控制在0.05水平上，但经过45次多重比较后，犯第一类错误的概率上升到：1-(1-0.05)45=0.9. 可见选用t检验法进行多重比较，仅仅控制了每次比较的显著水平，但却大大增加了整体的显著水平。

所要控制的几种误差类型和选用的检验方法：

（1）第一类误差率——即犯第一类错误的概率α;

（2）比较误差率——即每一次单独比较时，所犯第一类错误的概率，可使用T法、LSD法、DUNCAN法；

（3）试验误差率——即完成全部比较后，整体所犯第一类错误的概率；

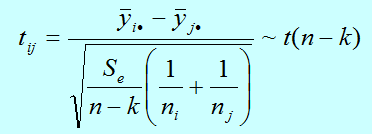
（4）完全无效假设下的试验误差率——即在H0假设完全无效下的试验误差率，可使用SNK法；

（5）部分无效假设下的试验误差率——即在H0假设部分无效下的试验误差率；

（6）最大试验误差率——即在在H0假设完全或部分无效下，完成全部比较后所犯第一类错误的最大概率，可使用BON法、SIDAK法、SCHEFFE法、TUKEY法、GT2/SMM法、GABRIEL法、REGWQ法、REGWF法、DUNNETT法。

1. T检验和Bonforroni检验

当因素有k个水平时，对任意两个水平均值间的差异的显著性检验，可用 t统计量：

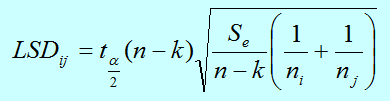


两两比较的次数共有，因此，共有m个置信水平，每次比较的显著水平取α，则完成所有比较后的整体显著水平为1-(1-α)m, 比较次数m越大，试验误差就越大。

而Bonforroni检验的方法取α/m. 完成所有比较后的整体显著水平等于1-(1-α/m)m<α.

2. LSD检验

既可以通过两两比较的显著水平的特定限制来控制最终的试验误差率，也可以通过两两比较的绝对差异界限来判别显著性。最容易想到的这个界限就是在两两比较中采用的*t*检验法而得到Fisher最小显著差（LSD）为



当时，则P值≤α.

3. **SNK检验和Duncan检验**

属于多级检验法，检验效率更高，分为步长增加法和步长减少法，SAS采用步长减少法。

当因素有k个水平时，即有k个均值需要比较，检验步骤为：

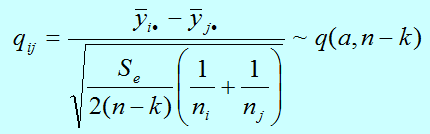
（1）将均值由大到小排列；

（2）比较第1个和第k个均值间（跨度a=k）是否有显著差异。若两者之间无显著差异，说明其他均值之差比它小的任何两个水平均值之间的差别也无显著性，所以停止一切比较；反之，则继续进行下一步；

（3）比较第1个和第k-1个、第2个与第k个均值（跨度a=k-1）是否有显著差异。若两者之间的比较无显著差异，则停止一切比较。否则，跨度减1，直到跨度为2为止。

多级检验在作每一级比较时，通过控制比较误差率γa的显著水平来实现其最终要控制的试验误差率。注意到γa是跨度和整体试验误差率的函数，它在每一级比较时可能是不同的。

实际上，γa就是每一级比较时特定统计量分布的显著水平。常用的两种方法是SNK检验和Duncan检验。它们的检验统计量为*q*（也称学生化极差统计量）：



中a是两均值之间的跨度值，q分布的自由度是a和n-k，显著水平为γa. SNK检验和Duncan检验的区别主要在于γa取值：

SNK检验：γa=α（注意，当比较次数很大时，最大试验误差率将趋向于1）；

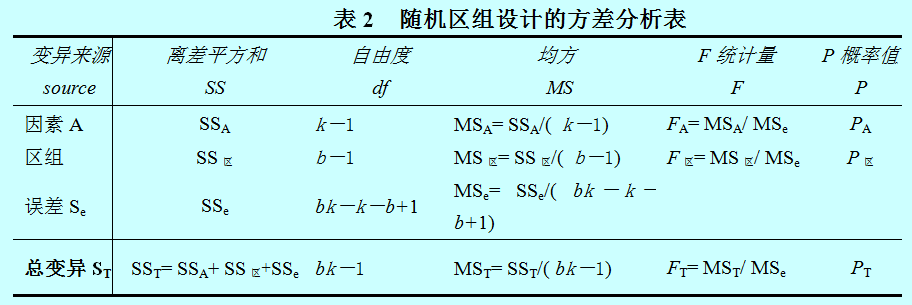
Duncan检验：γa=1-(1-α)a-1.

**六、随机区组设计的方差分析**

是两样本配对试验的扩大版，是要比较因素A中的k个水平的各个均值。试验设计时，先将受试对象按性质相同或相近者组成区组，每个区组有k个受试对象，分别随机分配到因素A的k个水平上。

这时每个水平的受试对象，不仅数量相同，而且性质也相同或相近，就能缩小误差，提高试验效率。这样的设计可将区组看作一个因素，就成为两个因素的设计（因素与区组），由于两个因素的各水平仅交叉1次，所以重复数为1，在这样的意义下，随机区组设计可看作为两因素重复数为1的设计，一般这种设计不考虑交互影响。

设因素A，k个水平，分成b个区组，每个区组有k个受试对象，分别随机分配到因素A的k个水平上。那么，随机区组设计的方差分析表为：



**七、析因设计的方差分析**

析因设计（Factorial Design）是一种多因素的设计。各因素在试验中所处的地位基本平等，而且因素之间存在一级（即2个因素之间）、二级（即3个因素之间）乃至更复杂的交互作用。

例如，两个因素时，第1个因素有3个水平，第2个因素有2个水平，全部水平组合共有3×2=6种组合，每种组合都作试验，就是析因试验设计，也可称为3×2析因试验设计。同样3×4×2析因试验设计，则代表3个因素，分别有3，4，2个水平，全部试验后的水平组合为3×4×2=24种。在每一种组合下，适当重复几次，称为重复数。重复数可以不相等，一般地说，重复数相等时，效率最高。

析因设计能够检验每个因素的各水平间主要变量的平均值的统计差异，也能检验因素间的交互影响。当存在交互影响时，表示一个因素各水平间的差异会随着另一个因素的水平改变而不同；当不存在交互影响时，则各个因素独立，即一个因素的水平改变时不影响另一个因素的各个水平之效应。

析因设计的方差分析因为能研究交互影响，所以能提供较多信息。但是，当有较高级（二级以上）的交互影响时，由于涉及多个因素，各有多个水平，情况将错综复杂，可能会引起解释上的困难。

析因设计的方差分析同样是从数据差异的总平方和开始分解。例如，对于*A*×*B*双因素方差分析，这个总差异能分解成4部分：

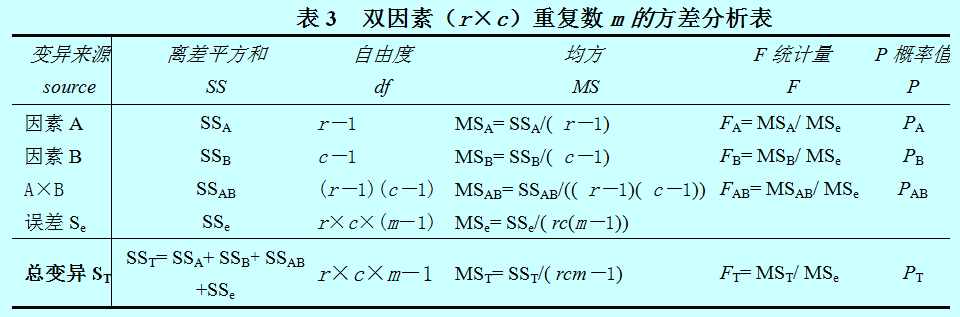
① A因素的各个水平之间的差异；

② B因素的各个水平之间的差异；

③ A与B的各种不同组合之间的差异；

④ 观察数据必然会产生的随机误差。

方差分析的主要目的就是要将这四部分从总平方和中分离出来，再以各个平方和与误差平方和作比较。假设A因素有r个水平，B因素有c个水平，每一种水平下的重复数为m，那么总的观察数据有n=r×c×m个，方差分析表如下：

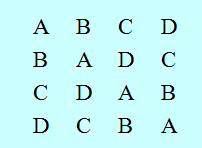


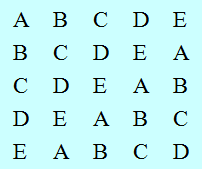
**八、拉丁方设计的方差分析**

若试验中涉及到3个因素，当它们之间不存在交互作用或交互作用可以忽略不计，且各因素均取相同水平时，适合于选择拉丁方设计。

用K个拉丁字母排成K行K列的方阵，使每行每列中每个字母仅出现1次，这样的方阵称为拉丁方（Latin Square）。然后将3个因素分别放置到拉丁方的行、列及字母上面。

4×4和5×5的拉丁方设计如下：





例如，三种中药组和对照组白兔血凝时间的4×4拉丁方设计：

