

第5章 方差分析

本章学习目标

- 学习并掌握方差分析概念,进行方差分析的基本条件。
- 学习并掌握利用 SAS 和 JMP 进行单因素方差分析的方法,理解输出统计量的意义及判断依据。
- 学习并掌握利用 SAS 进行多因素方差分析的方法,理解方差分析输出统计量的意义及判断依据。
- 学习并掌握利用 SAS 进行协方差分析的方法,理解协方差分析输出统计量的意义及判断依据。

5.1 方差分析简介

5.1.1 方差分析基本概念

方差分析是数理统计学中常用的数据分析方法之一,它是分析实验数据各因素各水平对某事物某指标的影响是否有显著差异的一种统计分析方法。根据实验考虑的因素个数及是否有协变量参与作用,可分为单因素、双因素、多因素方差分析及协方差分析。在方差分析中,经常把实验数据的总方差分解为由所考察因素引起的主因素方差、因素间的交互作用引起的交互因素方差和随机因素引起的随机误差方差。当影响因素是定性变量(一般称为分组变量或效应变量),观测结果是定量变量(一般称为结果变量或响应变量)时,常用的数据处理方法是:假设响应变量均数或均值相等,从而检验响应变量受因素影响的效果是否有显著差异。

若只有一个效应变量(也称因素或因子),而且其水平数 $K=2$,则称为单因素 2 水平方差分析,最常用 t 检验进行两两均值比较以检验对响应变量影响的效果是否产生显著差异。SAS 系统的 TTEST 过程即可完全满足此类数据分析的需要。若实验中考虑单因素多水平, $K>3$,或考虑两个或两个以上因素且多水平时($k \geq 2$),则称多因素多水平方差分析,通常用多重比较的 F 检验。SAS 系统的 ANOVA 和 GLM 过程可完全满足多因素多水平的多重比较的需要。

F 检验又叫方差齐性检验。从两个研究总体中随机抽取两组样本,要对这两组样本进行比较的时候,首先要判断两总体方差是否相同,即方差齐性,就要用 F 检验。F 检验就是检验两组样本的方差是否有显著性差异,以确定它们的精密度是否存在差异,也就是确定两组数据之间是否存在系统误差。当不存在系统误差时,可进行 t 检验。正态性和方差齐性是进行方差分析的两个基本条件。

当不能满足方差齐性的条件时,至今尚未找到十分满意的处理方法,仅能采用非参数检验法进行简单分析。尽管如此,由于方差分析适用的范围比较广泛,所以,它在假设检验中起到了举足轻重的作用。因此,弄清方差分析的基本思想,有助于读者尽快学会如何用方差分析法处理各种实验设计方案下的实验数据。

方差分析的基本思想:当实验数据满足正态性和方差相等(也称方差同质性)的假设条件时,把全部数据关于总均数的离差平方和分解成几部分,每一部分表示某一影响因素或诸影响因素之间的交互作用所产生的效应。将各部分均方(即方差)与误差均方相比较,依据F统计量做出统计推断,得到统计结论。

1. 单因素试验的方差分析

考虑一个因素A取 a 个水平,分析这 a 个不同水平对所考察的指标Y的影响,即在实验中只有A一种因素取 a 个水平变化,而其他因素控制不变,这样的实验叫单因素实验,所进行的方差分析叫单因素实验的方差分析。

零假设: $H_0: \mu_1 = \mu_2 = \cdots = \mu_a$

备择假设: $H_a: \mu_i \neq \mu_j$, 至少有一对这样的 i, j

在 A_i 水平下的样本均值为:

$$\bar{x}_{i.} = \frac{1}{n_i} \sum_{j=1}^{n_i} x_{ij}$$

样本数据的总平均值为:

$$\bar{x} = \frac{1}{n} \sum_{i=1}^a \sum_{j=1}^{n_i} x_{ij}$$

总离差平方和为:

$$SS_T = \sum_{i=1}^a \sum_{j=1}^{n_i} (x_{ij} - \bar{x})^2$$

总离差平方和可分解为:

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

其中,上式中的最后一项为0。

若记:因素A的离差平方和为:

$$SS_A = \sum_{i=1}^a \sum_{j=1}^{n_i} (\bar{x}_{i.} - \bar{x})^2$$

误差项的离差平方和为:

$$SS_e = \sum_{i=1}^a \sum_{j=1}^{n_i} (x_{ij} - \bar{x}_{i.})^2$$

则有: $SS_T = SS_A + SS_e$ (总变差 = 组间差异 + 组内差异)。

SS_T 是全部实验数据与总平均值之间的差异,称为总变差。 SS_A 表示在 A_i 水平下的样本均值与总平均值之间的差异,叫因素A效应的离差平方和,也称组间差异。 SS_e 表示在 A_i 水平下的样本值与该水平的样本均值之间的差异,它是由随机误差引起的,叫误差平方

和,又称组内差异。

SS_T 的自由度为 $n-1$ 。因为 SS_A 与 SS_e 相互独立, SS_A 的自由度 f_A 为 $a-1$ (a 为水平数), SS_e 的自由度 f_e 为 $n-a$ 。

$$f_T = f_A + f_e \quad F = \frac{\frac{SS_A}{f_A}}{\frac{SS_e}{f_e}}$$

2. 双因素无交互作用的方差分析

客观现实中的事物很复杂,影响某项指标的因素往往有很多,这些因素互相联系,互相依存,互相对立,问题也变得复杂多样。当只考虑两个因素的作用,且两因素间无交互作用时,我们进行组间变差和组内变差(即误差)的变差分析,叫双因素无交互作用实验方差分析。

双因素无交互作用总离差平方和分解为:

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

$$SS_T = SS_A + SS_B + SS_e$$

SS_T 的自由度为 $ab-1$ 。因为 SS_A 、 SS_B 与 SS_e 相互独立, SS_A 的自由度 f_A 为 $a-1$ (a 为 A 因素的水平数), SS_B 的自由度 f_B 为 $b-1$ (b 为 B 因素的水平数), SS_e 的自由度 f_e 为

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

$$F_1 = \frac{\frac{SS_A}{f_A}}{\frac{SS_e}{f_e}} \quad F_2 = \frac{\frac{SS_B}{f_B}}{\frac{SS_e}{f_e}}$$

3. 双因素有交互作用的方差分析

$$SS_T = SS_A + SS_B + SS_{A \times B} + SS_e$$

$$SS_e = SS_T - SS_A - SS_B - SS_{A \times B}$$

$$SS_A = bn \sum_{i=1}^a (\bar{x}_{i.} - \bar{x})^2$$

$$SS_B = an \sum_{j=1}^b (\bar{x}_{.j} - \bar{x})^2$$

$$SS_{A \times B} = n \sum_{i=1}^a \sum_{j=1}^b (\bar{x}_{ij.} - \bar{x}_{i.} - \bar{x}_{.j} + \bar{x})^2$$

$$SS_e = \sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^n (x_{ijk} - \bar{x}_{ij.})^2$$

SS_T 的自由度为 $abn-1$ 。因为 SS_A 、 SS_B 与 SS_e 相互独立, SS_A 的自由度 f_A 为 $a-1$ (a 为 A 因素的水平数), SS_B 的自由度 f_B 为 $b-1$ (b 为 B 因素的水平数), $SS_{A \times B}$ 的自由度 $f_{A \times B}$ 为 $(a-1)(b-1)$, SS_e 的自由度 f_e 为

$$(abn-1) - (a-1) - (b-1) - (a-1)(b-1) = ab(n-1)$$

$$F_1 = \frac{\frac{SS_A}{f_A}}{\frac{SS_e}{f_e}} \quad F_2 = \frac{\frac{SS_B}{f_B}}{\frac{SS_e}{f_e}} \quad F_{A \times B} = \frac{\frac{SS_{A \times B}}{f_{A \times B}}}{\frac{SS_e}{f_e}}$$

4. 多因素有交互作用的方差分析

方差分解

$$SS_T = SS_A + SS_B + SS_C + SS_{A \times B} + SS_{A \times C} + SS_{B \times C} + SS_{A \times B \times C} + SS_e$$

5. 多重比较

在三个或多个均值之间做两个或多个均值比较的检验称为多重比较 (Multiple Comparison Procedure)。

在引入假设概念时,曾把选择 α 水平与做出错误判断的风险相联系。为了解多重比较方法,需要控制所有比较的总错判的机会(当均值相同时判断为不同的错误),也要控制每个单独比较的错判机会,因此分为实验比较错误率和比较错误率两大类,实验比较错误率 MEER 相对比较错误率 CER 更精确、更严格。

例如,有 5 种施肥方法,均值做两两比较共有 10 种组合,如要控制 10 种比较的总错判率在 α 水平,平均到每组比较的控制水平会在十分之一的 α 水平,这就称为控制“实验比较错误率”或称“控制对整个实验的总错判机会”,记为“实验比较错误率”(Means Experimentwise Error Rate, MEER)。

另一种,如果想对 10 种比较中的每一种进行单独控制比较错判机会在 α 水平,就称为控制“比较错误率”,或称“对每一个比较的单独错判机会”,记为“比较错误率”(Comparisonwise Error Rate, CER)。以上例为例,每一种两两比较分别进行 α 水平 CER 的控制,则 10 种两两比较合计则为 10 倍的 α 水平,相对 MEER 而言则每种比较的错判率会扩大 10 倍。因此在多重比较时,采用 MEER 方法进行比较会比 CER 方法比较要严谨得多。

在进行多重比较时,应采用两类方法进行比较,若 MEER 方法能够得到较好的结论,一定以 MEER 法的结论作为研究结论。若 MEER 法比较均值无显著差异,可采用 CER 法进行比较,可用 CER 法的结论作为研究结论。

对于多因素,由于多因素的影响及各因素间的交互作用相当复杂,为了获得最好的实验结论,通常优先使用 MEER“实验比较错误率”作为控制准则,其次可选用 CER 比较错误率。

SAS 软件在 GLM 方差分析过程中提供了多种 MEER 法和多种 CER 法,两大类中的各种方法都有其特点,选用时应分清各种方法的优势,合理使用。

5.1.2 方差的同质性检验

1. 方差检验的基本思路

方差检验的基本思路是:利用样本方差建立一个统计量,并为这个总体方差的统计量构造一个置信区间,这个置信区间的显著水平为 α ,区间中包括总体方差的概率是 $1-\alpha$ 。

在确定的 α 水平下,统计量有其固定的拒绝区域。

单尾检验中,拒绝区域分布在统计量分布曲线的一侧;

双尾检验中,拒绝区域分布在统计量分布曲线的两侧。

如果检验统计量大于或等于临界值而落入拒绝区域,或 P 值小于显著水平 α 而落入拒绝区域,便可以拒绝零假设;反之,则不能拒绝零假设。

2. 方差同质性检验过程

(1) 提出零假设 H_0 和备择假设 H_a :

$$H_0: \sigma^2 = \sigma_0^2 \qquad H_a: \sigma^2 \neq \sigma_0^2$$

(2) 构造检验统计量:

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

在 H_0 成立的条件下,统计量 χ^2 服从自由度为 $n-1$ 的 χ^2 分布。

或构造 F 统计量: $F = \frac{s_1^2}{s_2^2}$ 在 H_0 成立的条件下,统计量 F 服从自由度为 n_1 和 n_2 的 F 分布。

(3) 确定显著水平 α (α 通常取 0.05)。

(4) 规定决策规则。

双尾检验时,拒绝区域在两侧,如果检验统计量大于右侧临界值或小于左侧临界值,则拒绝零假设。若是单尾检验,拒绝区域分布在一侧,具体选择左侧还是右侧,可根据备择假设 H_a 的情况而定。

(5) 进行判断决策。

所谓方差的同质性(也称方差齐性),就是指各个总体的方差是相同的。方差的同质性检验就是要从各样本的方差来推断其总体方差是否相同。

3. 两组样本方差的同质性检验

假设两组样本的样本容量分别为 n_1 和 n_2 ,方差分别是 s_1^2 和 s_2^2 (一般将方差数值较大的样本方差记为 s_1^2),总体方差分别为 σ_1^2 和 σ_2^2 。当检验两个总体方差 σ_1^2 和 σ_2^2 是否同质时,可用 F 检验法。当两样本所属总体服从正态分布,且两组样本的抽样是随机和独立的时,其 F 值等于两组样本方差 s_1^2 和 s_2^2 之比,即 $F = \frac{s_1^2}{s_2^2}$,并服从 $df_1 = n_1 - 1, df_2 = n_2 - 1$ 的 F 分布。

当 $F < F_\alpha$ 时,接受 $H_0: \sigma_1^2 = \sigma_2^2$,即认为两组样本方差同质;

当 $F > F_\alpha$ 时,否定 $H_0: \sigma_1^2 = \sigma_2^2$,接受 $H_a: \sigma_1^2 \neq \sigma_2^2$,即认为两组样本方差不同质。

$$\text{Var} = s^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2 = \frac{\text{SS}}{n-1} = \frac{\text{SS}}{\text{df}}$$

4. 多组样本方差的同质性检验

对三组或三组以上样本方差进行同质性检验,一般采用 Bartlett 检验方法。

假设 $H_0: \sigma_1^2 = \sigma_2^2 = \cdots = \sigma_k^2$,即 k 组样本的方差同质。

$H_a: \sigma_1^2 \neq \sigma_2^2 \cdots \neq \sigma_k^2$,不完全相等。

对 k 组独立样本方差, 求其合并方差 s_p^2 , 矫正数 C 和 χ^2 :

$$s_p^2 = \frac{\sum_{i=1}^k s_i^2 (n_i - 1)}{\sum_{i=1}^k (n_i - 1)}$$

$$C = 1 + \frac{1}{3(k-1)} \left[\sum_{i=1}^k \frac{1}{n_i - 1} - \frac{1}{\sum_{i=1}^k (n_i - 1)} \right]$$

$$\chi^2 = \frac{2.3026}{C} (\lg s_p^2 \sum_{i=1}^k (n_i - 1) - \sum_{i=1}^k (n_i - 1) \lg s_i^2)$$

上式服从 $df=k-1$ 的 χ^2 分布。其中, $2.3026=\ln 10$ 。

对确定的显著水平 α , 如果 $\chi^2 < \chi_0^2$, 接受 H_0 : 认为 $\sigma_1^2 = \sigma_2^2 = \cdots = \sigma_k^2$ 成立; 如果 $\chi^2 > \chi_0^2$, 则否定 H_0 , 接受 H_a : $\sigma_1^2 \neq \sigma_2^2 \cdots \neq \sigma_k^2$ (至少有一组方差不等), 表明这些样本方差不同质。

5.1.3 方差分析的基本假定和数据转换

1. 方差分析的基本假定

对实验数据进行方差分析前提条件是数据符合正态性和方差同质性, 如果分析的数据不符合这两条基本假定, 则不能采用方差分析。

(1) 正态性: 实验误差应是服从正态分布的独立的随机变量。因为方差分析只能估计随机误差, 顺序排列或顺序取样资料不能使用方差分析。正态分布要求每一个观测值 x_{ij} 应围绕其平均数呈正态分布, 非正态分布的原始数据资料经过适当的数据转换后, 若近似服从正态分布, 也可对转换的变量进行方差分析。

(2) 可加性: 处理效应与误差效应应该是可加的, 并服从方差分析的数学模型, 即 $x_{ij} = \mu + \alpha_i + \epsilon_{ij}$, 这样才能将实验的总变异分解为各因素引起的变异, 以确定各变异在总变异中所占的比例, 对实验结果做出客观评价。

(3) 方差同质性: 所有实验数据的方差应具备同质性, 也称方差齐性, 即 $\sigma_1^2 = \sigma_2^2 = \cdots = \sigma_n^2$ 。因为方差分析是将各个处理的实验误差合并以得到一个共同误差的方差, 所以必须假定资料中有这样一个共同方差存在。方差异质将使假设检验中某些处理效应得出不正确的结果。如果发现方差不同质的现象, 可将变异性特别明显的数据剔除。当然剔除数据时应十分小心, 以免失掉某些重要信息。或者将实验分成几个部分进行分析, 使每个部分满足方差同质性, 满足假定条件, 以进行方差分析。

2. 数据转换

有时遇到的样本, 其所来自的总体和上面提到的方差分析基本假定相抵触, 这些数据在做方差分析之前必须经过适当的处理, 如进行相应的数据转换。有时样本的非正态性、不可加性和方差的异质性通常连带出现, 这类数据分析时主要考虑处理效应与误差效应的可加性, 其次考虑方差的同质性。

常用的简单数据转换方法如下。

1) 平方根转换

有些观测数据为泊松分布而非正态分布,比如一定面积上某种杂草株数或昆虫头数等,样本平均数与其方差有比例关系,采用平方根转换可获得方差同质性。一般将原观测值转换成 $\sqrt{x}$,数据较小时采用 $\sqrt{x+1}$ 。

例如,表 5-1 中所列的一定面积燕麦田中某杂草的株数。从直观上看, A_1 、 A_2 和 A_3 、 A_4 及 A_5 间的数据相差太大,方差同质性不成立。

表 5-1 燕麦田中某种杂草的株数

处理	A_1	A_2	A_3	A_4	A_5
1	438.00	538.00	77.00	17.00	18.00
2	442.00	422.00	61.00	31.00	26.00
3	319.00	377.00	151.00	87.00	77.00
4	380.00	315.00	52.00	16.00	20.00
$\bar{x}_i$	395.00	413.00	87.00	38.00	35.00

对表 5-1 资料,如果计算出误差项的方差,它可能是个平均值,用以检验 A_1 、 A_2 间差数则太小,用来检验 A_3 、 A_4 和 A_5 间的差数则太大。如把表中数据进行平方根转换(如表 5-2 所示),可以看出各处理资料的范围就相差不多。对表 5-2 中数据做方差分析,列入表 5-3。

表 5-2 燕麦田中某种杂草株数的平方根

处理	A_1	A_2	A_3	A_4	A_5
1	20.90	23.20	8.80	4.10	4.20
2	21.00	20.50	7.80	5.60	5.10
3	17.90	19.40	12.30	9.30	8.80
4	19.50	17.70	7.20	4.00	4.50
$\bar{x}_i$	19.80	20.20	9.00	5.80	5.70

表 5-3 表 5-2 资料的方差分析表

变 异 来 源	df	ss	s^2	F	$F_{0.05}$	$F_{0.01}$
处理误差时间	4	866.663	216.666	46.425**	3.06	4.89
	15	69.995	4.667			
总变异	19	936.658				

2) 对数转换

如果已知资料中的效应成比例而不是可加的,或者标准差(或极差)与平均数大体成比例时,可以使用对数转换。

以下数据是在 5 次三个相继的夜里捕获昆虫的几何平均数,实验采用三种捕蛾灯诱捕(见表 5-4)。

表 5-4 捕获昆虫统计及捕获数的对数值

时期	捕蛾灯			对数值		
	I	II	III	I	II	III
1	19.1	50.1	123.0	1.28	1.70	2.09
2	23.4	166.0	407.4	1.37	2.22	2.61
3	39.5	223.9	398.1	1.60	2.35	2.60
4	23.4	58.9	229.1	1.37	1.77	2.36
5	16.6	64.6	251.2	1.22	1.81	2.40
$\bar{x}_i$	22.4	112.7	281.8	1.37	1.97	2.41
极差	22.9	173.8	284.4	0.38	0.65	0.52

原来的平均数和极差近于正比关系,经对数转换后三个极差较为接近,且与平均数无关。将转换后的数据进行方差分析(见表 5-5)。

表 5-5 对数转换后的数据的方差分析表

变异来源	df	SS	s^2	F	$F_{0.05}$	$F_{0.01}$
时期间捕蛾灯 误差	4	0.4941	0.1235	8.52*	3.84	7.01
	2	2.7503	1.3752	94.84**	4.46	8.65
	8	0.1159	0.0145			
总变异	14	3.3603				

用 LSD 法比较三种捕蛾灯之间的差异,证明不同捕蛾灯之间有显著差异。

3) 反正弦转换

如果数据是以比例数或以百分率表示的,其分布趋向于二项分布,方差分析时应做反正弦转换,用下面的公式把它们转换成一个相应的角度:

$$\theta = \sin^{-1} \sqrt{P}$$

式中, P 为百分数资料, θ 为相应的角度值。

5.2 ANOVA 过程和 GLM 过程简介

5.2.1 ANOVA 过程

SAS 软件有两个过程步: ANOVA 和 GLM 可进行各种方差分析,每个过程步通过不同的选项为用户提供了不同的均值比较方法。

ANOVA 过程主要用于处理均衡设计(即对于每个因素、每个水平的观测数是相等的,另外还可以处理拉丁方设计、正交设计等)的一元、多元方差分析,也可用于多个变量的对比检验。

实验设计如果不均衡,建议使用 GLM 过程。ANOVA 过程和 GLM 过程最后需用

QUIT 语句退出。

1. ANOVA 过程语句格式

```
PROC ANOVA 选择项 1 ;
  CLASS 变量 ;
  MODEL 因变量=效应变量/选择项 2 ;
  MANOVA H=效应变量 E=效应变量;
  BY 变量;
  MEANS 效应变量/选择项 3;
RUN; Quit;
```

2. 选择项 1

- (1) DATA=SAS 数据集名。
- (2) MANOVA: 要求 PROC ANOVA 按多元方式删除那些含有丢失值的观测,即只要在因变量中有丢失值,就在分析中删除这条观测。
- (3) OUTSTAT=SAS 数据集: 定义一个输出数据集,其中含有平方和,F 统计量以及模型中每个效应的概率水平。如果在 MANOVA 语句中指定了 CANONICAL 选择项,且没有 M= 选择项,数据集中还含有典型分析的一些结果。

3. CLASS 语句

在方差分析中,区分分类水平的变量叫分类变量,分类变量必须在 CLASS 语句中说明。CLASS 语句中的变量取值个数叫水平数。分类变量的取值可以是数值型,也可以是字符型,但所表示的意义都是不同水平,而因变量 Y 则必须是连续型的数值型变量。

4. MODEL 语句

MODEL 语句用来指明效应变量和响应变量间的关系。如果没有指明效应,则 ANOVA 只拟合截距项,也只检验效应变量是否为 0。

MODEL 语句斜杠(/)后的选择项 2:

- (1) INT | INTERCEPT: 要求 ANOVA 过程把截距项作为一个效应进行处理,打印出与其有关的假设检验结果。ANOVA 过程在模型拟合时总是含有截距,但是,当这个选项省略时,不输出与其有关的假设检验结果。当不考虑模型中的截距项的效应时,可以用 NOINT 选项去掉截距项。
- (2) NOINT: 不考虑模型中的截距项的效应。
- (3) NOUNI: 不输出单变量分析结果。

5. MANOVA 语句

如果 MODEL 语句中含有多于一个的响应变量,就可以使用 MANOVA 语句要求进行

多元方差分析。使用了 MANOVA 语句之后,如果某一个观测自变量或效应变量中有丢失值,则在方差分析中删除这条观测。

(1) **H=效应变量**: 指定模型中一些效应作为假设检验的矩阵。对于每一个 H 矩阵, H=选择项打印出 $E^{-1}H$ 的特征根及特征向量(其中, E 是和误差效应相关的矩阵)等。

(2) **E=效应变量**: 规定该变量为误差效应,如果省略,则使用误差 SSCP(残差)矩阵。

6. MEANS 语句中的选择项 3

用来计算 MEANS 语句中列出的每个效应所对应的因变量各水平的均值及均值间的显著性。

(1) **Hovtest**: 方差同质性检验。仅针对单因素进行方差同质性检验,或称等方差检验。

选择项 3 提供了多种均值比较的 MEER 法和 CER 法。

(2) **MEER 法有: BON、REGWQ、SNK、GABRIEL**。

① **BON**: 对于 MEANS 语句中的主效应均值之差进行 Bonferroni 的 t 检验,控制实验错误率 MEER,就是把每个检验的 α 水平取小一些进行多重 t 检验。对整个实验选一个 α 水平,然后除以两两比较的检验次数。如要进行 10 次两两检验,而将 MEER 控制在 0.05 水平,则每次检验需控制在 0.005 水平。

② **REGWQ**: 对于 MEANS 语句中的所有主效应进行多重 F 检验。REGWQ 检验控制 MEER。BON 和 t 检验都用到了比较每一对均值的检验,而 REGWQ 检验是一种多级检验。它先检验所有 N 个均值间的差异,再对 $N-1$ 个均值的各集合进行检验,再对 $N-2$ 个各集合进行检验,以此类推。

③ **SNK**: 对 MEANS 语句中指定的全部主效应执行 Student-Newman-Keuls 组间多重比较。

④ **GABRIEL**: 对于 MEANS 语句中的所有主效应进行 Gabriel 多重对比检验。

(3) **CER 法有: DONCAN、T|LSD、DUNNETT、DUNNETTL、DUNNETTU**。

① **DONCAN**: 对 MEANS 语句中指定的全部效应执行组间多重比较的 Duncan's 检验。

② **T | LSD**: 对 MEANS 语句中的所有主效应进行两两 t 检验,相当于在单元观测数相等时的 Fisher's 检验。用重复 t 检验控制 CER(比较错误率),每对做一次,控制每对检验的比较错误率 CER。

③ **DUNNETT**: 对 MEANS 语句中指定的全部效应执行各组与对照组间多重比较的双尾 Dunnett's t 检验。如果要指定对照组所对应的(分组)变量水平,将该水平所对应的格式化变量值以单引号括起置于选项后的圆括号中。如果要为多个主效应指定对照组,将各效应之对照组所对应的变量值(先以单引号括起)以空格分隔置于选项后的圆括号中。默认情况下,各效应的第一个水平将被作为对照组使用。

例如:

MEANS A/DUNNETT ('A1');

其中,'A1'是 A 效应的对照组取值。