临床医师实用统计学讲座

第六讲　多组均数的对比分析
方差分析法

皖南医学院　孙瑞元　黄志力

两样本均数的比较用t检验或u检验，而多个样本均数的比较则需用方差分析法，方差分析法又称F值法，其基本思想是把全部观察值之间的变异——总变异，按设计和需要分为二个或多个组成部分，再作分析。方差分析法效率高，性能全面，适合于以下情况的分析：（1）两组或多组均数间的显著性检验；（2）多列两组均数的对比分析；（3）多因素多水平对实验的影响；（4）分析组间变异、组内变异、次数间变异、标本间变异……等各种因素以及这些因素间的交互影响；（5）为拉丁方设计、正交设计及生物检定等实验进行析因分析……等。方差分析的计算较繁，过去普及不够，目前在计算器的帮助下，是值得重视及推广的一种统计方法。

一、单因素方差分析法

在医学研究中，常要比较不同病种的生理生化指标是否相同，某病治疗前后不同时间的生理生化指标有无变化，不同季节对水中某种物质的含量有无显著影响等问题。这里我们把研究的病种、时间、季节等称为因素。仅按不同病种，不同时间、不同季节的分组称为单因素分组。

以例1为例介绍传统的方差分析计算方法。

例1：某湖水不同季节氯化物含量测定值见表1(上半部分)所示，问不同季节氯化物含量有无差别？

表1的结果可以看出：（1）组内变异：每个季节内部的测定值不尽相同。显然它不是季节不同的影响，而只是由于个体差异和随机测量等所致。（2）组间变异：各季节的湖水氯化物含量的均数不相同。这表明不同季节对湖水氯化物含量可能有一定的影响，当然也包含有误差的作用。（3）总变异：31次测定的湖水氯化物含量，有高有低，这又是一种变异。它既可能受不同季节的影响，也包括误差的作用。

不同季节的湖水氯化物含量的均数之间的变异是由于不同季节的影响，还是误差所致呢？为了得出结论，宜采用方差分析。

1.总变异的离均差平方和(SS总)。即31个观察值与总均数($\bar{\bar{X}}$)之差的平方和。经数理推导得：

$$SS_{总}=\Sigma X^2-\frac{(\Sigma X)^2}{N}\qquad 式(1)$$

式(1)中：X表示各个测定值，N为总例数。ΣX^2是各观察值平方之和，$(\Sigma X)^2$是总计数的平方。本例$SS_{总}=11057.8-(577.6)^2/31=295.8$

表1　某湖水不同季节氯化物含量　(mg/L)

	春	夏	秋	冬	
	22.6	19.1	18.9	19.0	
	22.8	22.8	13.6	16.9	
	21.0	24.5	17.2	17.6	
	16.9	18.0	15.1	14.8	
	24.0	15.2	16.6	13.1	
	21.9	18.4	14.2	16.9	
	21.5	20.1	16.7	16.2	
	21.2	21.2	19.6		Σ
$\sum_{j=1}^{n_i} X_{ij}$	171.9	159.3	131.9	114.5	577.6(ΣX)
n_i	8	8	8	7	31(n)
$\bar{X}_i$	21.49	19.91	16.49	16.36	18.56($\bar{\bar{X}}$)
$\sum_{j=1}^{n_i} X^2_{ij}$	3724.51	3231.95	2206.27	1895.07	11057.8(ΣX^2)

2. 组间变异的离均差平方和（SS组间）。SS组间即各季节氯化物含量的均数($\bar{X}$)与总均数($\bar{\bar{X}}$)之差的平方和：

$$SS组间 = \Sigma n \cdot (\bar{X} - \bar{\bar{X}})^2 \quad 式(2)$$

式(2)中n为各组例数。

本例SS组间 $= 8(21.49-18.56)^2 + 8(19.91-18.56)^2 + 8(16.49-18.56)^2 + 7(16.36-18.56) = 151.3$

3. 组内变异的离均差平方和（SS组内）。SS组内即每个季节内部的观察值与其均数($\bar{X}$)之差的平方和。按此定义算得春、夏、秋、冬的值分别为30.8、59.9、31.6、22.2。故本例SS组内＝30.8＋59.9＋31.6＋22.2＝144.5

由于本例只做单因素分析，数理论证及上述结果均证明式(3)成立。

$$SS总 = SS组间 + SS组内 \quad 式(3)$$

所以只需算出SS总及SS组间，SS组内很易得出，本例SS组内＝295.8－151.3＝144.5和硬算结果相同（硬算时计算量较大）。

4. 均方及F值：为了便于对比，用组内均方（J组内）、组间均方（J组间）分别代表组内变异及组间变异，其大小可按式(4)、式(5)计算。

$$J组内 = \frac{SS组内}{n-K} \quad 式(4)$$

$$J组间 = \frac{SS组间}{K-1} \quad 式(5)$$

式中n为测定值的总个数，K为组数。

两均方的比值即为F值，见下式：

$$F = \frac{J大}{J小} \quad 式(6)$$

本例：$J组内 = \frac{144.5}{31-4} = 5.4$

$$J组间 = \frac{151.3}{4-1} = 50.4$$

$$F = \frac{50.4}{5.4} = 9.33$$

5. 结果判断：根据自由度f查F值表，如：$F \geq F_{0.05}$，则$P \leq 0.05$；$F \geq F_{0.01}$，则$P \leq 0.01$；$F < F_{0.05}$则$P > 0.05$。

本例组间自由度$f_1 = K-1 = 3$，组内自由度$f_2 = n-K = 31-4 = 27$，查F值表$F_{0.01}(3,27) = 4.60$。今$F = 9.33 > F_{0.01}(3,27)$。故$P < 0.01$，说明各季节湖水氯化物含量的差别有高度显著性，可认为不同季节的湖水氯化物含量是不相同的。

上述分析文字较多，是为了帮助同志们理解方差分析，在实际工作中只要按上述的公式算出SS组间、SS组内、SS总、J组间、J组内、F等值，根据自由度查F值表，进行结果判断。一般都列出方差分析表。本例方差分析表见表2。

表2 本例方差分析表

变异来源	SS	f	J	F
总变异	295.8	30		
组间变异	151.3	3	50.4	9.33
组内变异	144.5	27	5.4	

传统的方差分析法计算繁复，例2中介绍简化方差分析法。

二、双因素方差分析法

例2：某次动物实验比较用药组及对照组的量反应变化，由于条件限制每批实验只能同时进行12只动物的观察，第一次实验后虽然看到用药组药效强于对照组，两组t检验的结果是差异无显著意义（t＝1.514，P＞0.05），于是在不同月份又进行了第二次及第三次实验。实验结果见表3。显然，这种多次实验无法用t检验，只能采用方差分析。

表3 三次用药的药效比较

实验次数	用药组	对照组	小计值
1	7,6,8,2,4,6(33)	4,4,5,2,6,2(23)	56
2	4,6,8,4,6,2(30)	3,2,2,5,3,4(19)	49
3	5,8,6,3,5,6(33)	4,5,3,2,2,4(20)	53
小计值	96	62	158

例1和例2不同，例1仅按季节不同分组，这叫按单因素分组，而例2除按用药组及对照组分组外，还按不同月份进行了三批实验并分组观察，这称双因素分组，下面以例2为例介绍简化方差分析法。

1. 总变异的标准差：实测值7，6，8……2，2，4等36个数据，反映着整个实验中总的变异情况。把36个数据直接输入计算器后即可算得标准差S：

$S = 1.856$，$S^2 = 1.856^2 = 3.445$。

我们的简化算法，将S值直接算出并引入其他公式，不再计算SS总，f总，$\frac{(\Sigma X)^2}{N}$等中间计算值，大大简化了计算过程。

2. 各因素的变异：行小计中56、49、53三值的标准差（S＝3.512）反映着次间变异；列小计中96及62的标准差（S＝24.041）反映着组间变异，行小计是12个实验值的总和（小计内含例数为12），而列小计的内含

例数为18。为了对比，应当取方差的均值，称为“均方”，以J表示。由于内含例数是总例数除以小计的个数，故：

各因素之均方 $J=S^2\cdot K/N$ 式(7)

各因素之自由度 $f=K-1$

式中 S为小计的标准差，可用计算器直接算出，K为小计值的个数，N为总例数。

本例次间均方为 $1.028(=3.512^2\times3/36)$，组间均方为 $32.110(24.041^2\times2/36)$。计算尤为简便。

(1)当均方来自两个小计值时，更可简化为：

$J=D^2/N$ 式(8)

D是两小计值之差，N为总例数。例如：组间均方可计算为 $96-62=34$，$J_{组间}=34^2/36=32.11$ 结果和用式(7)算出的结果相同。这种小计值个数为2的情况在生物检定，正交实验、药效分析中颇为多见，故式(8)很有价值。

(2)前面的公式要求各小计值的内含例数必须相等，若不等，宜采用式(9)计算均方：

$$J=\left[\Sigma\frac{Z^2}{n}-\frac{(\Sigma X)^2}{N}\right]/(K-1) \quad 式(9)$$

Z为各小计值，n为该小计的内含例数。

在临床工作中，通过合理的实验设计，使内含例数相等并不困难，故式(9)并非常用。

3. 误差项的变异：在总变异中去除各因素的变异，残余的就称为实验误差或残余误差，它反映着实验中的个体差异或实验波动。传统的方差分析计算很麻烦，现用式(10)很易算出。

误差的均方 $J_e=[S^2(N-1)-\Sigma fJ]/f_e$ 式(10)

误差的自由度 $f_e=N-1-\Sigma f$

式中ΣfJ是各因素的均方与自由度之积的总和，Σf是各因素自由度之和。本例已算出，$S=1.856$，$N=36$，$\Sigma f=2+1=3$，$\Sigma fJ=2\times1.028+1\times32.11=34.166$，故$f_e=36-1-3=32$，$J_e=[1.856^2\times(36-1)-34.166]/32=2.699$。

4. F值显著性检验：比较各J值与Je值，如某因素均方(J)近于或小于误差均方(Je)，表示该因素的变异与实验误差波动相近或更小，显然，该因素引起的差异无显著意义；如J比Je相差倍数越大，则该因素也就越重要，J/Je的值即为F值。

$F=J/J_e$ 式(11)

根据f及fe查F值表，进行结果判断（标准同单因素分析中的标准)。

通常$F<1.5$，可认为$P>0.05$，$F>10$可肯定$P<0.05$，这样不查F值表即可作出判断。

5. 在资料分析中，常将上述计算结果列成表4。

表4 三次实验的方差分析表

变异来源	均方	自由度	F	P
组　间	32.11	1	11.897	<0.01
次　间	1.028	2	<1	>0.05
误　差	2.699	32	—	—

按式(10)可算出组间均方32.11，$F=32.11/2.699=11.897$，根据自由度$f=1$，$f_e=32$，查F值表得$F_{0.01}=7.50$，现$F>F_{0.01}$，$P<0.01$，故组间差异有极显著意义，说明用药组的药效变化大于对照组。

次间变异的均方为0.695，小于Je，故$F<1$，无显著意义$(P>0.05)$。

和传统的方差分析法相比，它省略了总变异及离均差平方和等值，这些数值既未提供更多信息，又未参与对比分析，故省略。

若用传统方法对例2进行分析，所得结果与简化法相同，但步骤很多，此处略，有趣者可自己计算。

三、讨论

1. 方差分析要求资料的分布为常态，且各组间方差相齐，和t检验相同，特殊资料可进行平方根、对数及倒数转换后再用此法。

2. 多组均数经F检验差异有显著意义时，不一定内部任两组均数间差异都有显著意义，应使用Q检验进行两两对比分析，若F检验无显著意义则无必要进行Q检验。Q检验请读者参阅有关书籍，本讲座不介绍Q检验。

3. 方差分析法用途很广，从实用的角度出发，除多组均数对比分析外，其它用途不再介绍。

简讯　全国首届雷公藤临床应用学术研讨会于1987年11月7日在湖北省洪湖市召开。来自全国108名长期从事雷公藤研究的专家教授出席了大会。大会收到论文113篇，分别在大会与分组会上进行了学术交流，并进行了热烈讨论，一致认为该药在治疗风湿性疾病、皮肤性疾病，具有显著疗效。对治疗肾脏疾病及其他疾病有较好的疗效。会议决定成立雷公藤研究的专业学组，组织进行全国性的协作攻关。尽早进行生药品种、资源调查的研究及统一药物质量标准。湖北省卫生厅副厅长、中国中西医结合研究会湖北分会理事长涂用宏出席了大会并讲了话。　　（李瑞琳）