

# 用 Excel 建模做基于样本统计量的单因素方差分析法

陈雄新

(湖南环境生物职业技术学院 护理学院,湖南 衡阳 421005)

**摘要:**生态学与医学领域的科研工作中经常需要对数据做方差分析,在已知样本例数、平均数、标准差等样本统计量的前提下,如果利用 Excel 软件制作方差分析模板,既可为数据分析提供方便快捷的方差分析手段,也可为审阅论文数据时,如何判断其结论的真实性提供了新方法.图 8,表 2,参 5.

**关键词:**单因素方差分析;样本统计量;Excel 建模;生态学研究;快捷分析与判断

**中图分类号:**R195.1

**文献标识码:**A

在生态学及医学领域的科研工作中经常需要对数据做方差分析,尽管统计软件有许多,但是多需要提供原始数据.当阅读科研论文时,经常遇到文中并没有提供原始数据仅列出了样本例数、平均数、标准差等样本统计量,而报告了方差分析的  $F$  值及其  $q$  检验结果的情况.那么,有没有一种方便快捷手段提供方差分析结果并判断其论文数据的真实性.因此,笔者介绍一种基于样本统计量并利用 Excel 建模做快速分析的单因素方差分析方法.

## 1 材料来源

用二氧化矽 50 mg 对大鼠染尘后,不同时期的全肺湿重变化如表 1<sup>[1]</sup>,试比较染尘后 1、3、6 月 3 个时期的湿重有无差别.

表 1 二氧化矽 50 mg 染尘后 3 个时期大鼠全肺湿重(单位:g)

Tab.1 Three periods of wet weights of rat whole lung exposed to 50 mg SiO<sub>2</sub> dust(unit:g)

1 月	3 月	6 月
3.3	4.4	3.6
3.6	4.4	4.4
4.3	3.4	5.1
4.1	4.2	5.0
4.2	4.7	5.5
3.3	4.2	4.7

某社区随机抽取了 30 名糖尿病患者、IGT 异常及正常人进行载脂蛋白(mg/dL)测定,比较 3 种人的载脂蛋白结果有无差别<sup>[2]</sup>,结果见表 2.

表 2 糖尿病患者、IGT 异常及正常人的载脂蛋白测定结果(单位:mg/dL)

Tab.2 Determination results of apolipoprotein in diabetics, IGT abnormal and normal persons(unit:mg/dL)

糖尿病	IGT 异常	正常人
85.7000	96.0000	144.0000
105.2000	124.5000	117.0000
109.5000	105.1000	110.0000
96.0000	76.4000	109.0000
115.2000	95.3000	103.0000
95.3000	110.0000	123.0000
110.0000	95.2000	127.0000
100.0000	99.0000	121.0000
125.6000	120.0000	159.0000
111.0000		115.0000
106.5000		

## 2 研究方法

### 2.1 单因素方差分析

在 Excel 中其操作步骤为:①打开一个 Excel 工作表的窗口,将例 1 的原始数据按下图的方式列出,见 A2-E8 的格子,见图 1.②单击 Excel 菜单“工具”

栏,→选择“数据分析”功能项,→选择“单因素方差分析”分析项,→单击“确定”按钮.③出现“单因素方差分析”对话框,在输入区域,选中包括原始数据 A2-C8 的所有格子,在输出区域,选中 \$I\$1,在标志位于第一行处打“√”,单击“确定”按钮.④在 \$I\$20 区域出现单因素方差分析结果,见图 2<sup>[3]</sup>.

	A	B	C	D	E
1	表1二氧化矽50mg染尘后三个时期大鼠全肺湿重				
2	1月	3月	6月		
3	3.3000	4.4000	3.6000		
4	3.6000	4.4000	4.4000		
5	4.3000	3.4000	5.1000		
6	4.1000	4.2000	5.0000		
7	4.2000	4.7000	5.5000		
8	3.3000	4.2000	4.7000		

图1 表1原始数据图

Fig.1 Original data chart of tab.1

I	J	K	L	M	N	O
方差分析: 单因素方差分析						
SUMMARY						
组	观测数	求和	平均	方差		
1月	6.0000	22.8000	3.8000	0.2080		
3月	6.0000	25.3000	4.2167	0.1937		
6月	6.0000	28.3000	4.7167	0.4377		
方差分析						
差异源	SS	df	MS	F	P-value	F crit
组间	2.5278	2.0000	1.2639	4.5175	0.0291	3.6823
组内	4.1967	15.0000	0.2798			
总计	6.7244	17.0000				

图2 表1单因素方差分析结果

Fig.2 Single factor analysis of variance results from tab.1

## 2.2 方差分析 Excel 模板的建立

对于完全随机设计的数据资料,一般使用原始数据进行单因素方差分析,如若给出了各处理组的平均数、标准差和样本例数,也可以进行单因素的方差分析.单因素方差分析参考公式如下:其中,  $MS_z$  为总均方,  $MS_{sj}$  为处理组的组间均方,  $MS_{wc}$  为误差均方,  $S_{cl}^2$  为各处理组数均数的方差,  $N$  为总例数,  $n$  为各处理组观察例数,  $k$  为处理组数,  $\bar{x}_i$  为第  $i$  个区组的均数,  $\bar{x}_j$  为第  $j$  个处理组的均数,  $v_z$  为总自由度,  $v_{sj}$  为处理组自由度,  $v_{wc}$  为误差自由度.

### 2.2.1 各比较组例数相等的单因素方差分析

如文献[1]的已知各组的观察例数、均数、标准差和总例数及比较组数分别为  $n$ ,  $\bar{x}_i$ ,  $s_i$ ,  $N$ ,  $k$ .

组间的均方为:

$$MS_{sj} = n \cdot s^2, s^2 = \frac{\sum \bar{X}_i^2 - (\sum \bar{X}_i)^2 / k}{k - 1},$$

其对应的自由度为  $v_{sj} = k - 1$ .

组内的均方为:

$$MS_{zn} = \frac{\sum s_i^2}{k},$$

其对应的自由度为  $v_{zn} = N - k$ ,  $F = MS_{sj} / MS_{zn}$ .

### 2.2.2 各比较组例数不相等的单因素方差分析

在文献[1]中如已知各组的观察例数、均数、标准差和总例数及比较组数、分别为  $n_i$ ,  $\bar{x}_i$ ,  $s_i$ ,  $N$ ,  $k$ . 则各组间均数的方差为  $s^2$ :

$$s^2 = \frac{\sum n_i \bar{X}_i^2 - (\sum n_i \bar{X}_i)^2 / N}{N - 1}$$

$$MS_{sj} = \frac{\sum n_i \bar{X}_i^2 - (\sum n_i \bar{X}_i)^2 / N}{k - 1} \times (N - 1)$$

$$= \frac{S^2(N - 1)}{k - 1}$$

$$MS_{zn} = \frac{\sum S_i^2(n_i - 1)}{N - k}$$

$$F = \frac{S^2(N - 1)(N - k)}{(k - 1) \sum S_i^2(n_i - 1)}$$

### 2.2.3 各比较组例数相等方差分析 Excel 模板编制

按下列公式编制模板:

E12 = SUM(B12:D12), B15 = B14^2, C15 = C14^2, D15 = D14^2,

B17 = (((B13)^2 + (C13)^2 + (D13)^2) - ((B13+C13+D13)^2/3)/(3-1),

B18 = B12 \* B17, B19 = B16 - 1, B20 = (B15 + C15 + D15)/B16, B21 = (B12 + C12 + D12) - B16, B22 = B18/B20.

分析模板见图 3.

	A	B	C	D	E
9	基于统计量的方差分析模板				
10	单因素方差分析--各比较组例数相等				
11	指标	1月	3月	6月	总观测数
12	观测数	6.0000	6.0000	6.0000	18.0000
13	平均数	3.8000	4.2167	4.7167	
14	标准差	0.4561	0.4401	0.6616	
15	各组方差	0.2080	0.1937	0.4377	
16	比较组数	3.0000			
17	总方差 $S^2$	0.2106			
18	组间MS	1.2639			
19	组间自由度	2.0000			
20	组内MS	0.2798			
21	组内自由度	15.0000			
22	F值	4.5175			

图3 基于表1的方差分析 Excel 模板(样本例数相等)

Fig.3 Variance analysis template by Excel based on tab.1 (sample cases are equal)

### 2.2.4 各比较组例数不相等方差分析 Excel 模板编制

按下列公式编制模板:

E16 = SUM(B16:D16), B19 = B18^2, C19 = C18^2,

2, D19 = D18^2,

$C21 = (((((B17)^2) * B16 + (C17^2) * C16 + (D17^2) * D16)) - (((B17 * B16 + C17 * C16 + D17 * D16)^2) / E16)) / (E16 - 1)$ ,  $C22 = C21 * (E16 - 1) / (B20 - 1)$ ,  $C23 = B20 - 1$ ,

$C24 = ((B18)^2 * (B16 - 1) + (C18)^2 * (C16 - 1) + (D18)^2 * (D16 - 1)) / (E16 - B20)$ ,

$C25 = E16 - B20$ ,  $C26 = C22 / C24$ .

分析模板见图 4.

	A	B	C	D	E
14	基于统计量的方差分析模板(单因素方差分析—各比较组例数不相等)				
15	指标	糖尿病	IGT异常	正常人	总观测数
16	观测数	11.0000	9.0000	10.0000	30.0000
17	平均	105.4545	102.3889	122.8000	
18	标准差	10.8731	14.5516	17.0672	
19	方差	118.2247	211.7486	291.2889	
20	比较组数	3.0000			
21	方差SS		82.2078		
22	组间MS		1192.0128		
23	自由度		2.0000		
24	组内MS		203.6236		
25	自由度		27.0000		
26	F值		5.8540		

图4 基于表2的方差分析 Excel 模板(样本例数不等)

Fig.4 Variance analysis template by Excel based on tab.2 (sample cases are unequal)

## 2.3 Excel 模板保存与应用

按上述方法可以列出更多组比较的数据,将模板适当命名后按 Excel 工作表文件格式保存,以便今后使用.使用时只需要将待检验的样本统计量(样本例数、平均数和标准差)数据代入图 3 模板内区域 B12:D14 或图 4 模板内区域 B16:D18 替换原来的相应数据,从模板中立即就可得到数据更新后的方差分析结果.

## 2.4 多个样本均数间的两两比较 $q$ 检验

### 2.4.1 样本例数相等的多个样本均数间的两两比较 $q$ 检验结果

按下列公式编制模板:

$C26 = D13 - B13$ ,  $C27 = C13 - B13$ ,  $C28 = D13 - C13$ ,  $D26 = \text{SQRT}((B20/2) * (1/B12 + 1/D12))$ ,  $D27 = \text{SQRT}((B20/2) * (1/C12 + 1/B12))$ ,  $D28 = \text{SQRT}((B20/2) * (1/C12 + 1/D12))$ ,  $E26 = C26 / D26$ ,  $E27 = C27 / D27$ ,  $E28 = C28 / D28$ . 结果见图 5.

三个样本均数间的两两比较 $q$ 检验法						
对比组	组数 $a$	均数 $\bar{x}_A - \bar{x}_B$	标准误 $s_{\bar{x}_A - \bar{x}_B}$	$q$ 值	$q$ 界值	P 值
A与B					$q_{0.05}$ $q_{0.01}$	
1与3	3.0000	0.9167	0.2160	4.2449	3.6700 4.8300	<0.05
1与2	2.0000	0.4167	0.2160	1.9296	3.0100 4.1700	>0.05
2与3	2.0000	0.5000	0.2160	2.3153	3.0100 4.1700	>0.05

图5 基于表1的两两比较  $q$  检验结果(样本例数相等)

Fig.5  $q$  test results by two comparisons based on tab.1 (sample cases are equal)

### 2.4.2 样本例数不等的多个样本均数间的两两比较 $q$ 检验结果

按下列公式编制模板:

$C30 = B17 - C17$ ,  $C31 = B17 - C17$ ,  $C32 = D17 - C17$ ,  $D30 = \text{SQRT}((C24/2) * (1/B16 + 1/D16))$ ,

$D31 = \text{SQRT}((C24/2) * (1/C16 + 1/B16))$ ,  $D32 = \text{SQRT}((C24/2) * (1/C16 + 1/D16))$ ,

$E30 = C30 / D30$ ,  $E31 = C31 / D31$ ,  $E32 = C32 / D32$ . 结果见图 6.

三个样本均数间的两两比较 $q$ 检验法						
对比组	组数 $a$	均数 $\bar{x}_A - \bar{x}_B$	标准误 $s_{\bar{x}_A - \bar{x}_B}$	$q$ 值	$q$ 界值	P 值
A与B					$q_{0.05}$ $q_{0.01}$	
1与3	3.0000	17.3455	4.4087	3.9344	3.4900 4.4500	<0.05
1与2	2.0000	3.0656	4.5352	0.6760	2.8900 3.8900	>0.05
2与3	2.0000	20.4111	4.6361	4.4026	2.8900 3.8900	<0.05

图6 基于表2的两两比较  $q$  检验结果(样本例数不等)

Fig.6  $q$  test results by two comparisons based on tab.2 (sample cases are unequal)

## 3 结 果

### 3.1 样本例数相等方差分析结果的验证

在图 3.基于表 1 的方差分析模板(样本例数相等)的单元格 B12:D14 位置分别代入需要检验的样本统计量(样本例数、平均数和标准差)数据后,马上出现方差分析的结果,见图 7.根据组间自由度和组内自由度查  $F$  界值表,与  $\alpha$  水平比较,可以作出统计学的结论.

	A	B	C	D	E
55	结果验证示例一:				
56	方差分析的模板(单因素方差分析—各比较组例数相等)				
57	指标	对照	模型	干预	总观测数
58	观测数	8.0000	8.0000	8.0000	24.0000
59	平均	10.0132	12.1364	11.2678	
60	标准差	0.9124	2.2661	1.7803	
61	方差	0.8325	5.1352	3.1695	
62	比较组数	3.0000			
63	方差SS	0.7926			
64	组间MS	9.1153			
65	自由度	2.0000			
66	组内MS	3.0457	F 界值		P 值
67	自由度	21.0000	0.0500	0.0100	
68	F 值	2.9928	3.4000	5.6100	>0.05

图7 样本例数相等方差分析结果的验证

Fig.7 Verification to results of variance analysis of equal sample cases

### 3.2 样本例数不等方差分析结果的验证

在图 4.基于表 2 的方差分析模板(样本例数不等)的单元格 B16:D18 位置,分别代入需要检验的样本统计量(样本例数、平均数和标准差)数据后,马上出现方差分析的结果,见图 8.根据组间自由度和组内自由度查  $F$  界值表,与  $\alpha$  水平比较,可以作

出统计学的结论

	A	B	C	D	E
70	结果验证示例二:				
71	(单因素方差分析—各比较组例数不等)				
72	指标	对照	模型	干预	总观测数
73	观测数	8.0000	5.0000	6.0000	19.0000
74	平均	3.0400	5.1300	4.2900	
75	标准差	0.3100	0.8700	0.7900	
76	方差	0.0961	0.7569	0.6241	
77	比较组数	3.0000			
78	方差SS	0.7921			
79	组间MS	7.1287			
80	自由度	2.0000			
81	组内MS	0.4263	F界值		P值
82	自由度	16.0000	0.0500	0.0100	
83	P值	16.7223	3.6300	6.2300	<0.05

图 8 样本例数不等方差分析结果的验证

Fig.8 Verification to results of variance analysis of unequal sample cases

4 结 论

通过两个实例,具体地说明了 Excel 建模在生态科学与医学领域科研中经常使用的完全随机设计的单因素方差分析的运用.其应用前提是使用者必须对统计学的基本概念和基本方法有所了解,并且对软件 Excel 的基本功能也有一定的了解.学会使用这一方法后,尤其是学会编制方差分析模板后,将为科研节约许多宝贵时间;并且,它可以帮助迅速判断科研论文中仅仅提供样本统计量所做的方差

分析结果的可靠性;而且,Excel 软件还可以快速和批量地做 $\chi^2$  检验<sup>[4]</sup>, $t$  检验<sup>[5]</sup>、 $u$  检验和相关分析等等;它可以与 Word、SAS、PowerPoint、FoxPro 和 SPSS 等软件之间进行数据的交换,使之发挥更多的作用.

参考文献:

[1] 马 燕.卫生统计学[M]. 北京:人民卫生出版社,2000.  
Ma Yan. Health Statistics [M]. Beijing: People's Medical Publishing House,2000.

[2] 倪宗赞.卫生统计学[M].北京:人民卫生出版社,2001.  
Ni Zong-zan. Health Statistics [M]. Beijing: People's Medical Publishing House,2001.

[3] 陈雄新.利用 Excel 软件做统计描述和方差分析的方法[J]. 湖南环境生物职业技术学院学报,2008,14(3): 17-19.  
Chen Xiong-xin.By using excel software to do statistical description and analysis of variance method [J].Journal of Hunan Polytechnic of Environment and Biology, 2008, 14 (3):17-19.

[4] 陈雄新.Excel 在  $t$  检验中的实用技巧[J].实用预防医学,2006,13(1):202-205.  
Chen Xiong-xin.Excel in t-test and practical skills [J]. Practical Preventive Medicine,2006,13(1):202-205.

[5] 陈雄新.Excel 在四格表 $\chi^2$  检验中的应用技巧[J].现代预防医学,2006,33(10):1 667-1 670.  
Chen Xiong-xin.The application of excel in four table chi-square test techniques [J].Modern Preventive Medicine, 2006,33(10):1 667-1 670.

Using Excel Modeling to Make Single Factor Analysis of Variance Based on Sample Statistics

CHEN Xiong-xin

( College of Nursing,Hunan Polytechnic of Environment and Biology,Hengyang 421005,China)

**Abstract:**In the scientific research of ecology or medicine,we often need to analyze the variance of the data.Under the premise that sample cases,average,standard deviation and other sample statistics are known,if using Excel software to produce variance analysis template,it can provide researchers with convenient and efficient means of variance analysis as well as a new method for us to judge the authenticity of its conclusions while reviewing the paper data.8figs.,2tabs.,5refs.

**Keywords:**single factor analysis of variance; sample statistics; Excel modeling; ecological studies; quick analysis and judgment

**Biography:**CHEN Xiong-xin,born in 1964,male,professor,research directions: medical statistics and epidemiology.