

用 Excel 建模做基于样本统计量的双因素方差分析法

陈雄新

(湖南环境生物职业技术学院 医学院,湖南 衡阳 421005)

摘要:在生态学与医学领域的科研工作中经常需要对实验数据做双因素的方差分析,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,可为研究者提供方差分析结果的方便快捷手段,同时,也为我们在审阅论文数据时,如何判断其结论的真实性提供了新方法.图 7,表 3,参 5.

关键词:双因素方差分析;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

在生态学与医学领域的科研工作中我们经常需要对实验数据做方差分析,尽管统计软件有许多,但是多需要提供原始数据.当我们在阅读科研论文时,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?笔者介绍一种基于样本统计量并利用 Excel 建模做双因素方差分析快速方法.

1 资料来源

为探讨 Rgl 对镉诱导大鼠睾丸损伤的保护作用,研究者按照窝别把大鼠分成 10 个区组,然后将同一区组内的 3 只大鼠随机地分配到 3 个实验组,分别给予不同处理,一定时间后测量大鼠的睾丸的 MT 含量($\mu\text{g/g}$),数据如表 1 所示^[1].试比较三种不同处理对大鼠 MT 含量有无差别.

表 1 三组大鼠 MT 含量值(单位: $\mu\text{g/g}$)

Tab.1 Three groups of rats' MT content value (unit: $\mu\text{g/g}$)

窝别	对照组	氯化镉组	Rgl+氯化镉组
1	40.6	78.3	116.3
2	44.8	86.0	124.6
3	36.7	72.1	149.0
4	49.9	95.4	128.8
5	59.8	99.2	134.1
6	54.5	95.9	133.0
7	38.4	76.4	115.6
8	41.6	79.9	117.0
9	46.8	86.5	128.4
10	44.7	85.3	124.3

2 研究方法

2.1 双因素方差分析

打开一个 Excel 工作表的窗口,将原始数据按下图的方式列出,见 A1-D11 的格子,见图 1.三组大鼠 MT 含量值的原始数据截图.

在 Excel 中其操作步骤为:①单击 Excel 菜单

窝别	对照组	氯化镉组	Rg1+氯化镉组
1	40.6	78.3	116.3
2	44.8	86	124.6
3	36.7	72.1	149
4	49.9	95.4	128.8
5	59.8	99.2	134.1
6	54.5	95.9	133
7	38.4	76.4	115.6
8	41.6	79.9	117
9	46.8	86.5	128.4
10	44.7	85.3	124.3

图 1 表 1 三组大鼠 MT 含量值的原始数据截图

Fig.1 Original data screenshot of three groups of rats' MT content value in tab.1

“工具”栏,→选择“数据分析”功能项,→②选择“方差分析:无重复双因素分析”分析项,→单击“确定”按钮;③出现“无重复双因素分析”对话框,在“输入区域”,选中包括原始数据 A1-D11 的所有格子,在标志位于第一行处打“√”,在“输出区域”,可任意选择某一区域,如果选中 \$ A \$ 16,单击“确定”按钮;④在被任意选中的 \$ A \$ 16 区域出现无重复双因素分析结果,见图 2^[2].

方差分析: 无重复双因素分析						
SUMMARY	观测数	求和	平均	方差		
1	3	235.2	78.4	1432.6		
2	3	255.4	85.1	1592.6		
3	3	257.8	85.9	3296.3		
4	3	274.1	91.4	1568.5		
5	3	293.1	97.7	1381.8		
6	3	283.4	94.5	1542.1		
7	3	230.4	76.8	1490.1		
8	3	238.5	79.5	1421.4		
9	3	261.7	87.2	1665.0		
10	3	254.3	84.8	1584.3		
对照组	10	457.8	45.8	52.7		
氯化镉组	10	855.0	85.5	82.3		
Rg1+氯化镉组	10	1271.1	127.1	103.7		
方差分析						
差异源	SS	df	MS	F	P-value	F crit
行	1277.0	9	141.9	2.93	0.024855	2.46
列	33078.8	2	16539.4	341.92	4.8E-15	3.55
误差	870.7	18	48.4			
总计	35226.5	29				

图 2 表 1 三组大鼠 MT 含量值的双因素分析结果图

Fig. 2 Two-factor variance analysis result of three groups of rats' MT content value in tab.1

2.2 双因素方差分析 Excel 模板的编制

2.2.1 无重复双因素分析公式

对于随机区组设计的数据资料,一般使用原始数据进行双因素的方差分析,若仅给出各处理组的平均数、标准差,而未给出各随机区组的平均数和

标准差时,无法完成双因素的方差分析,当给出各处理组、各随机区组的平均数和标准差时,则可以进行双因素的方差分析^[3].参考公式如下:

$$F_{cl} = \frac{MS_{cl}}{MS_{wc}}$$

$$F_b = \frac{MS_b}{MS_{WC}}$$

$$S_{cl}^2 = \frac{\sum X_i^2 - (\sum \bar{X}_i)^2/k}{k-1},$$

$$S_b^2 = \frac{\sum X_j^2 - (\sum \bar{X}_j)^2/n}{n-1}$$

$$MS_{cl} = n \cdot s_{cl}^2,$$

$$MS_b = k \cdot s_b^2,$$

$$MS_{wc} = (\sum s_i^2 - MS_b)/(k-1)$$

$$\nu_z = k \cdot n - 1,$$

$$\nu_{cl} = k - 1,$$

$$\nu_b = k - 1,$$

$$\nu_{wc} = \nu_z - \nu_{cl} - \nu_b$$

其中, MS_{cl} 为处理组的均方, MS_b 为区组的均方, MS_{wc} 为误差的均方. S_{cl}^2 为各处理组数平均数的方差, S_b^2 为各区组平均数的方差. N 为总例数, n 为区组数, k 为处理组数. x_i 为第 i 个区组的平均数, x_j 为第 j 个处理组的平均数. ν_z 为总自由度, ν_{cl} 为处理组自由度, ν_b 为区组自由度, ν_{wc} 为误差自由度.

2.2.2 “无重复双因素方差分析”模板的编制

第一步: 输入样本统计量. 在 \$A\$ \$18-\$ \$F\$ \$35\$ 区域输入各处理组的观察数、平均数、标准差; 输入各随机区组的观察数、平均数、标准差; 输入总例数、总平均数、总标准差; 输入比较组数 k 、区组数 b . 见图 3.

SUMMARY	观测数	求和	平均	方差	标准差
1	3	235.2	78.4	1432.6	37.9
2	3	255.4	85.1	1592.6	39.9
3	3	257.8	85.9	3296.3	57.4
4	3	274.1	91.4	1568.5	39.6
5	3	293.1	97.7	1381.8	37.2
6	3	283.4	94.5	1542.1	39.3
7	3	230.4	76.8	1490.1	38.6
8	3	238.5	79.5	1421.4	37.7
9	3	261.7	87.2	1665.0	40.8
10	3	254.3	84.8	1584.3	39.8
合计	30	2583.9	86.1	16974.8	130.3
对照组	10.0	457.8	45.8	52.7	7.3
氯化镉组	10.0	855	85.5	82.3	9.1
Rg1+氯化镉组	10.0	1271.1	127.1	103.7	10.2
比较组数k	3.0				
区组数b	10.0				
总例数N	30.0				

图 3 各处理组、各随机区组及合计的观察数、平均数、标准差的模板图

Fig.3 Template figure of every treatment group, every randomized block group and the total of observation, the mean and standard deviation

第二步: 建立“无重复双因素方差分析”的 Excel 模板. 按照方差分析参考公式在 \$A\$ \$36-\$ \$F\$ \$41\$ 区域建立 Excel 分析模板, 在 \$B\$ \$37\$、\$B\$ \$38\$、\$D\$ \$36\$、\$D\$ \$37\$、\$D\$ \$38\$、\$D\$ \$39\$、\$F\$ \$36\$、\$F\$ \$37\$、\$F\$ \$38\$、\$F\$ \$39\$、\$B\$ \$40\$、\$B\$ \$41\$、\$D\$ \$40\$、\$D\$ \$41\$、\$F\$ \$40\$、\$F\$ \$41\$ 格子内按照以下公式编制: 出现图 4 的分析结果, 作为模板保存以备今后使用.

$$S_{cl}^2 = ((D30^2 + D31^2 + D32^2) - (D30 + D31 + D32)^2/B33)/(B33-1),$$

$$S_b^2 = ((D19^2 + D20^2 + D21^2 + D22^2 + D23^2 + D24^2 + D25^2 + D26^2 + D27^2 + D28^2) - (D29 * B34)^2/B34)/(B34-1)$$

$$\nu_z = B33 * B34 - 1,$$

$$\nu_{cl} = B33 - 1,$$

$$\nu_b = B34 - 1,$$

$$\nu_e = D36 - D37 - D38$$

$$MS_z = F37 + F38 + F39,$$

$$MS_{cl} = B37 * B34,$$

$$MS_b = B38 * B33,$$

$$MS_e = (F30^2 + F31^2 + F32^2 - F38)/(B33 - 1)$$

$$F_{cl} = F37/F39, F_b = F38/F39,$$

$$F_{0.05} \text{ 界值}(cl) = \text{FINV}(0.05, D37, D39),$$

$$F_{0.05} \text{ 界值}(b) = \text{FINV}(0.05, D38, D39),$$

$$P \text{ 值}(cl) = \text{FDIST}(B41, D38, D39),$$

$$P \text{ 值}(b) = \text{FDIST}(B40, D37, D39).$$

双因素方差分析模板		$V_{\text{总}}$	29.0	$MS_{\text{总}}$	16729.7
$S^2_{\text{总}}$	1653.9	$V_{\text{组}}$	2.0	$MS_{\text{组}}$	16539.4
S^2_{b}	47.3	V_{b}	9.0	MS_{b}	141.9
		V_{e}	18.0	MS_{e}	48.4
$F_{\text{总}}$	341.9	$F_{0.05(V_{\text{总}}, V_{\text{e}})}$	3.5546	P值	0.0000
F_{b}	2.9	$F_{0.05(V_{\text{b}}, V_{\text{e}})}$	2.4563	P值	0.0249

图 4 双因素方差分析模板截图

Fig.4 Template screenshot of two-factor variance analysis

第三步:建立“多个样本均数间的两两比较 q 检验”的 Excel 模板.在 \$ A \$ 43- \$ H \$ 47 区域按下
列公式编制模板,见图 5.

$$\begin{aligned} \$ C \$ 45 &= E32-E30, \\ \$ C \$ 46 &= E31-E30, \\ \$ C \$ 47 &= E32-E31, \end{aligned}$$

$$\begin{aligned} \$ D \$ 45 &= \text{SQRT}((F39/2) * (1/B30+1/B32)), \\ \$ D \$ 46 &= \text{SQRT}((F39/2) * (1/B30+1/B31)), \\ \$ D \$ 47 &= \text{SQRT}((F39/2) * (1/B31+1/B32)), \\ \$ E \$ 45 &= C45/D45, \\ \$ E \$ 46 &= C46/D46, \\ \$ E \$ 47 &= C47/D47. \end{aligned}$$

三个样本均数间的两两比较 q 检验法							
对比组 A与B	组数a	均数	标准误	q值	q界值		P值
		$\bar{x}_A - \bar{x}_B$	$s_{\bar{x}_A - \bar{x}_B}$		0.0500	0.0100	
1与3	3.0000	51.0037	2.1994	23.1901	3.6100	4.7000	<0.01
1与2	2.0000	29.6404	2.1994	13.4768	2.9700	4.0700	<0.01
2与3	2.0000	21.3632	2.1994	9.7133	2.9700	4.0700	<0.01

图 5 无重复双因素方差分析两两比较 q 检验模板截图

Fig.5 Template screenshot of Q test by comparing each other

2.2.3 “无重复双因素分析”模板的保存与应用

按上述方法可以列出更多组比较的数据,将模板适当命名后将 Excel 工作表文件格式保存,以便今后使用.下次在使用时只需要将待检验的样本统计量(各处理组、各随机区组及合计的观察数、平均数、标准差的模板样本例数、平均数和标准差)数据代入图 3.各处理组、各随机区组及合计的观察数、平均数、标准差的模板中 \$ A \$ 18- \$ F \$ 35 区域以替换原来的相应数据,从图 4 双因素方差分析模板中就可立即得到数据更新后的方差分析结果.与此同时,从图 5 无重复双因素方差分析两两比较 q 检验模板中可得到相应的两两比较 q 检验结果.

3 结 果

3.1 双因素方差分析与两两比较 q 检验结果的验证

如某厂医务室测定了 10 名氟作业工人在工前、工中、工后 4 小时三个不同时间的尿氟浓度(μmol/L),问氟作业工人在这三个不同时间的尿氟浓度有无差别^[3],见表 2.

表 2 氟作业工人不同时间的尿氟浓度(单位:μmol/L)

Tab.2 Urinary fluoride concentration in different time (unit:μmol/L)among fluorine workers

工人编号	工前	工中	工后
1	90.53	142.12	87.38
2	88.43	163.17	65.27
3	47.37	63.16	68.43
4	175.80	166.33	210.54
5	100.01	144.75	194.75
6	46.32	126.33	65.27
7	73.69	138.96	200.02
8	105.27	126.33	100.01
9	86.32	121.06	105.27
10	60.01	73.69	58.95

经统计描述得出 10 名氟作业工人、在三个不同时间的尿氟浓度的样本统计量数据结果见表 3.

将表 3 的 10 名氟作业工人、在三个不同时间的尿氟浓度的样本统计量数据中的各处理组、各随机区组和合计的观察数、平均数、标准差数据分别代入已经保存好的图 3.各处理组、各随机区组及合计的观察数、平均数、标准差的模板图之单元格 \$ B \$ 19- \$ B \$ 35、\$ D \$ 19- \$ D \$ 32、\$ F \$ 19- \$ F \$ 32 位置处,马上出现双因素方差分析与两两比较 q 检验结果,见图 6 和图 7.

表 3 10 名氟作业工人、在三个不同时间的尿氟浓度的样本统计量数据
Tab.3 Sample statistics data from 10 fluorine workers in three different time of urinary fluoride concentration

SUMMARY	观察数	求和	平均	方差	标准差
1	3	320.03	106.68	944.65	30.74
2	3	316.87	105.62	2617.81	51.16
3	3	178.96	59.65	120.10	10.96
4	3	552.67	184.22	541.85	23.28
5	3	439.51	146.50	2246.22	47.39
6	3	237.92	79.31	1748.17	41.81
7	3	412.67	137.56	3991.29	63.18
8	3	331.61	110.54	193.99	13.93
9	3	312.65	104.22	302.55	17.39
10	3	192.65	64.22	67.59	8.22
合计	30	3295.54	109.85	115172.56	339.37
工前	10	873.75	87.38	1392.56	37.32
工中	10	1265.90	126.59	1165.98	34.15
工后	10	1155.89	115.59	3783.52	61.51

双因素方差分析模板		V_z	29.0	MS_z	9468.8
S^2_{c1}	409.14	V_{c1}	2.0	MS_{c1}	4091.4
S^2_b	1470.85	v_b	9.0	MS_b	4412.6
		v_e	18.0	MS_e	964.8
F_{c1}	4.24	$F_{0.05}$ 界值	3.555	P值	0.003
F_b	4.57	$F_{0.05}$ 界值	2.456	P值	0.031

图 6 10 名氟作业工人、在三个不同时间的尿氟浓度双因素方差分析结果
Fig.6 Results of two-factor variance analysis from 10 fluorine workers
in three different time of urinary fluoride concentration

三个样本均数间的两两比较q检验法						
对比组	组数a	均数	标准误	q值	q界值	
		$\bar{x}_A - \bar{x}_B$	$S_{\bar{x}_A - \bar{x}_B}$		0.0500	0.0100
A与B						
1与3	3.0000	28.2140	9.8222	2.8725	3.6100	4.7000
1与2	2.0000	39.2150	9.8222	3.9925	2.9700	4.0700
2与3	2.0000	11.0010	9.8222	1.1200	2.9700	4.0700

图 7 10 名氟作业工人、在三个不同时间的尿氟浓度两两比较 q 检验结果
Fig.7 Results of Q test by comparing each other from 10 fluorine workers
in three different time of urinary fluoride concentration

4 结 论

本文通过两个实例论证并说明了在生态科学与医学领域科研中基于样本统计量之下的随机区

组设计双因素方差分析的 Excel 建模运用结果.其应用前提使用者必须对统计学基本概念和基本方法以及 Excel 的基本功能都有所了解.当我们学会利用 Excel 公式编辑功能的这一方法后,尤其是学会自己编制单因素与双因素分析方差分析的模板后,

将为我们科研中节约时间;更为重要的是它还可以帮助我们对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做出迅速判断;利用 Excel 软件还可以快速和批量地做 χ^2 检验^[4], t 检验^[5]、 u 检验和相关分析等;而且利用 Excel 软件它可以与 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] 陈雄新.医学统计学[M].北京:北京出版社,2015.
Chen Xiong-Xin, editor. Health Statistics [M]. Beijing: Beijing House, 2015.
- [2] 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, (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, (3): 17-19.
- [3] 贺石林,陈修.医学科研方法导论[M].北京:人民卫生出版社 1998.
He Shi-Lin, Chen Xiu. editor. [M]. Beijing: People's Medical Publishing House, 1998.
- [4] 陈雄新.Excel 在 t 检验中的实用技巧[J].实用预防医学, 2006, (1): 202-205.
Chen Xiong-Xin. Excel in Test and Practical Skills [J]. Practical Preventive Medicine, 2006, (1): 202-205.
- [5] 陈雄新.Excel 在四格表 χ^2 检验中的应用技巧[J].现代预防医学, 2006, (10): 1 667-1 670.
Chen Xiong-Xin. The Application of Excel in Four Table Chi-square Test Techniques [J]. Modern Preventive Medicine, 2006, (10): 1 667-1 670.

Using Excel Modeling to Make Two-Factor Variance Analysis Based on Sample Statistics

CHEN Xiong-xin

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

Abstract: In the scientific research of ecology or medicine, we often need to do two-factor variance analysis of the experimental 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. 7figs., 3tabs., 5refs.

Keywords: two-factor variance analysis; sample statistics; Excel modeling; ecological studies; quick analysis and judgment.

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