

Microsoft Excel 方差分析的使用^{*}

郭文久

(云南农业大学农业科学技术学院, 昆明 650201)

摘要: 在介绍了 Excel 数据分析功能的安装后, 描述了如何用 Excel 进行农业试验设计的方差分析。Excel 的单因素方差分析方法可以进行单因素完全随机试验设计的方差分析; Excel 的无重复方差分析方法可以进行单因素随机区组和无重复的双因素方差分析; Excel 的可重复双因素方差分析方法可以进行系统分组资料和完全随机的双因素试验设计的方差分析。

关键词: Excel; 方差分析; 农业试验设计与统计分析

中图分类号: S 11⁺4.1 **文献标识码:** A **文章编号:** 1004-390X(2000)01-0009-04

在农业院校用的农业试验设计与统计分析的计算机计算中, 传统做法较多是采用 DOS 操作系统上提供的 BASIC、高级 BASIC 和 True BASIC 等开发的应用程序。但随着计算机产业的发展, 微型机的 CPU 由 8088 一直发展到 Pentium III, 自从微机 CPU 进入到 Pentium 特别是 Pentium 100 MHz 之后, 微机几乎全是安装的 WINDOWS 95 或者更高如 WINDOWS 98 等操作系统, 不能简单地把老的 DOS 版本的 BASIC 程序拿到现在的 WINDOWS 95 这类操作系统上来运行。因此必须寻找新的计算机计算方法。虽然现在也有象 SPSSWIN 这样优秀的在 WINDOWS 95 上用的统计软件, 但其未汉化, 不易使用。

现在介绍一种在 Microsoft Excel 上有关统计计算的方法。Microsoft Excel 本身有大量的内部函数和加载宏, 同时 Excel 还有强大的可编程能力, 具有几乎无限制的计算能力, 所有的农业试验的统计数据都可以计算。这种计算方法简单易行。并且将 Excel 与 Word 相结合, 同时进行数据计算和图表文章的编排, 就可以写出外观和内容都非常精美的文章来。而这些都是不需要太多时间的学习, 因为 WINDOWS 95 本身就是一种易学易用的图形操作系统。

1 应用准备“数据分析”功能的安装

Excel 的统计分析功能在 Excel 5.0, 6.0, 7.0 及 97 版中几乎是一致的, 变动较少, 但有的版本中这部分没有汉化, 而在 Excel 7.0 版中全部都是中文汉化过的, 易于使用, 这里介绍的内容就是以 Excel 7.0 版为准进行说明。

启动 Excel 7.0 后查看窗口主菜单“工具”项下最后一项(一般是最后一项)是否是“数据分析...”菜单项。若有表明已经安装了统计分析功能。若没有此项则应通过 setup 安装程序添加这部分的安装。在光驱放入 Poffice 95 光盘, 运行 Poffice 95 的 setup.exe 程序。见到“安装 Microsoft Office 95...”窗口时, 鼠标左键单击“添加/删除”按钮, 进入下一个窗口时, 在左边的选项框中选中“Microsoft Excel”项。鼠标左键单击使此项变黑即为选中。单击右边的“更改选项”按钮, 进入下一个窗口后, 单击左边选项框中的“加载宏”复选框, 使此项左边的小方框中有“√”符号, 单击“确定”按钮, 回到前一窗口后, 再单击“继续”按钮即进入了“数据分析”功能的安装过程。待安装完毕后再次启动 Microsoft Excel 7.0, 单击窗口主菜单的“工具”项, 再单击其下的“加载宏...”项, 在“加载宏”对话框中, 在左边的选项框中选中“分析工具库”和“分析工具

* 收稿日期: 1999-06-08

作者简介: 郭文久(1964-), 男, 重庆市人, 云南农业大学讲师, 主要从事生物统计研究。

库-VBA 函数”两项,使这两项左边的小方框中有“√”符号,单击“确定”按钮。再次单击“工具”菜单后,在最后一项上就有了“数据分析...”菜单项了,表明统计分析的内容已经安装完毕。

2 Excel 的方差分析

在窗口的主菜单上单击“工具”菜单,再选取“数据分析...”子菜单项即出现“数据分析”对话框选择左边的列表框中的分析项目,再单击“确定”按钮,即进入了相应的分析程序。注意在下面的论述中,输入数据表都以 Excel 工作表 sheet 的形式表示,而输出表没有采用此方式。

2.1 单因素方差分析 单因素方差分析仅适用于完全随机设计的单因素方差分析,以参考文献^[1]第 99 页为例说明 Excel 如何进行单因素方差分析,数据见表 1。在主菜单“工具”下选“数据分析...”,再在“数据分析”对话框中选择“单因素方差分析”项后又出现“单因素方差分析”对话框。

其中:“数据区域”指参加计算数据的 Excel 引用区域,此引用中可以包括 1 列标题,按表 1 此引用应为 \$A \$2: \$E \$6,格式为此区域的开始行列号加上英文的冒号(;)后再输入区域结束的行列号, \$表示绝对应用。其实也无须键入,在选中此编辑框后用鼠标在 Excel 的 sheet 上按住左键拖动,所显示的虚线框即输入了引用区域,如有错误,重新拖选即可。其实,在 Excel 中凡是要求输入引用区域的地方都可照此办理。

表 1 水稻盆栽试验的产量结果表

Tab. 1 Rice yields in potted cultivation experiment

	A	B	C	D	E
1			处理		
2	(1)	24	30	28	26
3	(2)	27	24	21	26
4	(3)	31	28	25	30
5	(4)	32	33	33	28
6	(5)	21	22	16	21

“分组方式”是指因素各水平的分组是按行还是按列排列的。本列是按行排列的所以应选择“行”。

“分类轴标记在第一列上”是指“数据区域”的第一列是否为标题。本例的第一列是标题,为因素各水平名称即(1)、(2)、(3)、(4)、(5),应选中此复

选框。

“ α ”是指在进行 F 测验时的显著概率水平,本例中 $\alpha = 0.05$ 。

“输出选项”分别指输出到“输出区域”指当前工作表的指定区域,如本例中输出到 \$A \$10 开始的区域,“新工作表”指输出到新的一张工作表中,在其后的编辑框中输入或点取工作表名,“新工作簿”指输出到新的 Excel 文件中。

在上面的对话框输入完毕后单击“确定”按钮,即出现表 2 的结果。

表 2 单因素方差分析结果表

Tab. 2 One factor ANOVA results

汇总				
组	计数	求和	平均	方差
(1)	4	108	27	6.66667
(2)	4	98	24.5	7
(3)	4	114	28.5	7
(4)	4	126	31.5	5.66667
(5)	4	80	20	7.33333

方差分析

差异源	SS	df	MS	F	P-值	临界值
组间	301.2	4	75.3	11.18317	0.00021	3.05557
组内	101	15	6.73333			
总计	402.2	19				

表 2 中的“汇总”表部分显示的是对每个水平的汇总结果。即依次排列的是水平名、水平的样本容量、总和、平均值和组内的方差,后面介绍的几种方差分析方法中也有“汇总表”,与此形式相同,限于篇幅未予列出。

在方差分析表中大部分项目我们都知道,这里只解释两项:“P-值”指 $F = 11.18317$, $df_1 = 4$, $df_2 = 15$ 时 F 分布的概率,可以用 Excel 的 =FDIST(F 值, 大均方自由度, 小均方自由度)函数取得进行验证;“临界值”指 $\alpha = 0.05$, $df_1 = 4$, $df_2 = 15$ 时 F 分布 F 的临界值,可以用 =FINV(显著水平, 大均方自由度, 小均方自由度)函数取得。这里我们可以根据“P-值”和“临界值”进行显著性判断,如“P-值”小于显著水平概率 α 值如 0.05, 则 F 检验显著,如“P-值”小于 0.01 则 F 检验达极显著,反之亦然;如实际 F 值大于“临界值”,则 F 检验显著,反之亦如此,但此显著性判断也仅限于在输入对话框中显著概率水平上显著或不显著,因此“临界值”不能得到象“P-值”那么多的信息。

表 3 烤烟品种产量比较试验结果表

Tab 3 Yields comparison experiment among varieties of tobacco

	A	B	C	D	E
1	区组	I	II	III	IV
2	A	15.3	14.9	16.2	16.2
3	B	18	17.6	18	18.3
4	C	16.6	17.8	17.6	17.8
5	D(CK)	16.4	17.3	17.3	17.8
6	E	13.7	13.6	13.9	14
7	F	17	17.6	18.2	17.5

表 4 无重复双因素方差分析结果表

Tab 4 Two factor ANOVA without repetition

变源	SS	df	MS	F	P-值	临界值
行	50.66333	5	10.13267	74.14146	4.9E-10	2.90130
列	2.325	3	0.77500	5.67073	0.00842	3.28738
误差	2.05	15	0.13667			
总计	55.03833	23				

2.2 方差分析:无重复的双因素分析 无重复的方差分析可以适用于双因素无重复的方差分析,这在农业试验中不存在。无重复的方差分析还可以适用于单因素随机区组的方差分析,这里用参考文献^[1]第 118 页的烤烟品种产量比较试验结果表为例说明 Excel 如何用此法进行单因素随机区组方差分析,资料见表 3。在“无重复双因素方差分析”对话框中,在“输入区域”中输入或选取 \$A \$1:\$E \$7,因包括行、列标记所以应选中“标记”复选框,其它同前,单击“确定”按钮。计算结果见表 4。

表 6 系统分组资料的方差分析表

Tab 6 ANOVA of system data blocks

变源	SS	df	MS	F	P-值	临界值
样本	251.2	2	125.6	96.61538	1.48E-17	3.19072
列	0.33333	3	0.11111	0.08547	0.96766	2.79806
交互作用	2.66667	6	0.44444	0.34188	0.91115	2.29460
组内	62.4	48	1.3			
总计	316.6	59				

注意数据的排列与原始资料上已有所不同。可重复的双因素分析资料的重复的排列应在行上。在“可重复的双因素方差分析”对话框中,“输入区域”应输入或选取 \$A \$2:\$E \$17,“每一样本的行数”指的是亚组内的重复数,这里为 5。注意如果

结果表中“行”为品种效应,“列”为区组效应。

2.3 方差分析:可重复的双因素分析 可重复的双因素分析,在我们农业试验设计中可以进行两种试验设计的统计分析:即系统分组资料(仅适用于“组内分亚组”的情况)和完全随机设计的双因素资料。现举例说明:

2.3.1 系统分组资料的方差分析

以参考文献^[1]第 107 页(例 6.5)的数据如表 5 为例进行说明:

表 5 3 种浓度的保水剂蘸根的发根数

Tab 5 Rooting numbers with roots treated with water conservative agent in three kinds of concentration

	A	B	C	D	E
1			盆	号	
2	保水剂浓度(i)	1	2	3	4
3	A ₁	8	7	8	9
4		7	7	8	7
5		8	9	7	8
6		7	8	9	7
7		7	8	8	7
8	A ₂	10	11	10	10
9		10	12	9	11
10		11	9	9	11
11		11	10	12	9
12		12	10	11	12
13	A ₃	13	14	13	12
14		12	15	13	14
15		12	12	14	12
16		11	11	14	13
17		14	11	10	14

这里输入错误计算结果将有所不同。其余同前。

输出的方差分析表如表 6,表中“样本”为主因素既“保水剂浓度间”的效应行,将“列”和“交互作用”行的 SS 和自由度加到一起既为“浓度内盆间”的亚组效应 SS_a和自由度。应该注意到系统分组

资料中组与亚组之间不应有交互作用;“组内”为误差 SS_{e2} 。在此基础上用 Excel 的计算公式和函数计算“浓度内盆间”的 MS、P-值和临界值。这里 $MS_i = 125.6, df_i = 2$, 则 $MS_{el} = (0.3333 + 2.6667)/(3+6) = 0.3333, df_{e1} = (3+6) = 9$, 则“保水剂浓度间”的 $F = 125.6/0.3333 = 380.7$, 浓度内盆间差异 $F = 0.3333/1.3 = 0.25$ 。“P-值”须用 Excel 提供的统计函数 FDIST 重新计算 F 分布的概率, 在单元格中输入 =FDIST(380.70, 2, 9), 显示结果为 2.01311E-09, 而临界值也应用 FINV 重新计算过, 输入 =FINV(0.05, 2, 9), 显示结果为 4.2565。对于其它的 F 分布概率和临界值也可以同样计算, 无须查阅教材上的统计表。

表 7 玉米品种与施肥二因素随机区组

Tab 7 Random blocks of two factors about maize varieties and fertilizers

	A	B	C	D	E
1			B 因素		
2	区组	A 因素	B1	B2	B3
3	I	A1	17	11	12
4	II		15	14	8
5	III		13	13	8
6	I	A2	19	20	17
7	II		13	19	16
8	III		11	13	18
9	I	A3	19	10	9
10	II		18	8	8
11	III		16	10	7

表 8 完全随机设计的双因素的方差分析表

Tab 8 Two factors ANOVA with completely random design

变源	SS	df	MS	F	P-值	临界值
样本	108.96296	2	54.48148	10.00680	0.00120	3.55456
列	81.40741	2	40.70370	7.47619	0.00433	3.55456
交互	148.14815	4	37.03704	6.80272	0.00161	2.92775
组内	98	18	5.44444			
总计	436.51852	26				

2.3.2 双因素完全随机设计的方差分析 我们用参考文献^[1]第 130 页的玉米品种与施肥二因素随机区组试验资料如表 7 为例说明双因素的方差分析, 注意资料的排列形式。这里不讨论怎样计算区组效应。对于表 7, 在可重复双因素方差分析对话框中, “输入区域”应为 \$B\$ \$2 : \$E \$ \$11, “每样本的行数”应为 3, 这里区组列不参加计算。计算结果如表 8。这里“变源”中“样本”为 A 因素效应、“列”为 B 因素效应、“交互作用”为 A×B 效应、“组内”即误差, 应该注意到这里的“组内”包括误差和区组效应在里面, 其实区组效应本来就是一种误差, 只不过是系统误差而已。

3 小结

在 Excel 中进行方差分析简便快捷, 但因为 Excel 统计部分的加载宏并非专为农业试验统计而写, 因此没有关于田间设计的计算方法。如农业试验设计中广泛使用的裂区试验设计和随机区组等

都没有计算方法。作者将在后面的文章中进行讨论。另外 Excel 的方差分析方法中还没有多重比较部分。

在本文中指出了 Excel 的 3 种方差分析方法所能计算的资料, 分别是: Excel 的单因素方差分析方法可以进行单因素完全随机设计的方差分析; Excel 的无重复双因素方差分析方法可以进行单因素随机区组和无重复的双因素方差分析; Excel 的可重复双因素方差分析方法可以进行系统分组资料和完全随机设计的双因素方差分析。Excel 的方差分析方法仅能计算到方差分析表为止。

参 考 文 献

- 1 四川农业大学, 西南农业大学, 云南农业大学主编. 农业试验与统计分析[M]. 成都: 四川科学技术出版社, 1993. 99, 107~108, 118~119, 130
- 2 Dodge, M. 著, 方中等译. Excel 5 for Windows 使用指南[M]. 北京: 清华大学出版社, 1994. 335~364

(下接 31 页)

Trail on Resistance of the Weed to Metsulfuron-methyl Herbicide

Wang Lin¹ Mei Hong¹ Nai Zhu²

(1 Institute of Plant Protection, YAAS, Kunming 650205)

(2 Biotechnology Research Institute of YAAS, Kunming 650223)

Abstract In this paper change of weed species and dominant population and amount change and height change of *Bidens pilosa* L. was reported after spraying herbicide Metsulfuron-methyl in survey plot of herbicide resistance plot during 1997~1998. Amount change from few to even more. Height change which becomes shorter with increase of treatment the drug concentration. The weed species and perennial weed were increased.

Key words Survey plot of herbicide resistance; Metsulfuron-methyl; Weed species; Dominant population; *Bidens pilosa* L.; Resistant change

(上接 12 页)

Variance Analysis by Microsoft Excel

Guo Wenjiu

(Faculty of Agricultural Science and Technology, Y A U, Kunming 650201)

Abstract After introducing the installation of Excel's Data Analysis Function, described three methods of analysis and how to use them for variance analysis of agricultural experiment data. Excel's one-factor variance analysis method can be used to one-factor variance analysis of complete randomized experiment design data. Excel's two-factor variance analysis method without repetition can be used to variance analysis of one-factor data with randomized blocks and two-factor data without duplication. Excel's two-factor variance analysis with duplication can be used to variance analysis of systematic block data and complete randomized two-factor experiment design data.

Key words Excel; Variance analysis; Agricultural experiment design and statistics analysis