Excel 在一年多点区域试验分析中的应用

韩永亮, 路正营, 李世云, 杨玉枫, 崔红印　(邯郸市农业科学院,河北邯郸 056001)

摘要　简述了 Excel 在一年多点区域试验分析中的应用。依据联合方差分析原理,介绍仅用 Excel 数据分析功能也可完成联合方差分析的新方法,其计算结果可靠。该研究为区域试验分析提供参考依据。

关键词　Excel; 方差分析; 区域试验;线性混合模型
中图分类号　S11　　**文献标识码**　A　　**文章编号**　0517 - 6611(2015)07 - 007 - 03

Application of Excel in Multi-location One Year Regional Experiments

HAN Yong-liang, LU Zheng-ying, LI Shi-yun et al　(Handan Academy of Agricultural Sciences, Handan, Hebei 056001)

Abstract　The application of Excel in multi-location one year regional experiments was elaborated. Based on the principle of variance analysis, the new method of multi-location one year experiments data analysis was introduced with Excel, which was basically in accordance with those from the DPS software. The study can provide reference for extend and application of the Excel software

Key words　Excel; Variance analysis; Regional crop trials; Linear mixed model

品种区域试验是由多个不同环境下的试验构成的通过这些试验分析和评定新品系在不同生态区域的丰产性、稳产性以及其他综合表现[1],是农作物品种审定的重要依据。准确、快速地分析试验结果是非常重要的。现已有多种专业性很强的统计分析软件如 SPSS、SAS、DPS 等[2]。这些软件虽然功能强大,但使用一般较复杂,需要一定时间的学习才能掌握[3]。在没有专用软件的况下,使用 Excel 2003 ~ 2007 进行分析,也可以完成区试的多点试验方差分析。笔者以区试 10 个品种、3 次重复、11 个试点为例,介绍 Excel 2007 的一年多点联合方差分析的操作方法。

1　一年多点试验基本原理

在品种区试中,多点试验起到了以空间换时间的作用。一年多点试验可以看作一个双因素随机区组试验[4]。设有 v 个品种,在 u 个地点做比较试验。每个地点皆设 r 个区组(重复),按随机区组设计进行试验,则第 i 个品种($i = 1, 2\cdots v$)在第 j 个地点($j = 1, 2\cdots u$)第 k 区组($k = 1, 2\cdots r$)的观测值为 $xijk$。它的线性模型为:

$$xijk = \mu + \tau i + vj + (\tau v)ij + \rho jk + \varepsilon ijk$$

式中,μ 为群体的平均值;τi 为品种 i 的效应值;vj 为地点 j 的效应。$(\tau v)ij$ 为品种 × 地点互作效应,ρjk 为地点内的区组效应,εijk 为随机误差[5-6],由此可以得到一年多点区域试验的方差分析表(表 1)。若方差分析中各效应的差异在 0.05 水平显著,则可进一步进行多重比较,以检查不同品种间或不同地区间的各效应是否存在差异。多重比较采用 Duncan's 新复极差方法,为此要计算出各效应差数的标准误。通过表 1 可以看出平均和(SS)的计算量通常较大,如果已知平方和,那么自由度和均方的计算非常简单,利用 Excel 公式就可以计算。因此,数据分析主要是平方和的分解。

表 1　一年多地试验的方差分析

变异来源	自由度(df)	平方和(SS)	均方(MS)	Ems(固定模型)
地点内区组	$u(r-1)$	SS_r	MS_r	
地点	$u-1$	SS_u	MS_u	$\sigma^2\varepsilon + r\nu\chi_u^2$
品种	$v-1$	SS_v	MS_v	$\sigma^2\varepsilon + r\nu\chi_v^2$
品种 × 地点	$(u-1)(v-1)$	SS_{vu}	MS_{vu}	$\sigma^2\varepsilon + r\chi_{uv}^2$
试验误差	$u(r-1)(v-1)$	SS_ε	MS_ε	$\sigma^2\varepsilon$
总计	$ruv-1$	SS_t		

2　基本方法

2.1　数据输入　该文的数据源自潘家驹主编的《作物育种学总论》[7]1981 年江苏省中籼稻区域试验数据(表 2)。数据分析软件为 Microsoft Excel 2007。P1 至 P10 分别代表 10 个品种,A1 至 A11 代表 11 个试点。按表 2 样式输入 Excel 工作表中即可。

2.2　单试点方差分析　在 Excel 工作表中完成数据录入后,开始对每个试点进行方差分析,以 A1 试点为例进行方差分析。首先,单击 Excel 菜单“数据”栏,选择“数据分析”功能项,再在对话框内选择“无重复双因素方差分析”分析项(图 1),下一步单击确定按钮出现下图的“无重复双因素方差分析”对话框(图 2);在输入区域,选中包括 A1 试点的原始数;在输出区域指定单元格,单击“确定”可以得到表 3 方差分析结果。

按照上面的方法,依次计算出 A2 至 A11 试点的方差分析表,其他试点的方差分析表省略,依据各点方差分析表写

作者简介　韩永亮(1979 -),男,河北磁县人,助理研究员,硕士,从事棉花遗传育种与生物统计学方面的研究。
收稿日期　2015-01-23

表 2 1981 年江苏省中籼稻区域试验数据

指标	p1	p2	p3	p4	p5	p6	p7	p8	p9	p_{10}
A1	25.15	24.73	24.03	24.51	24.43	23.90	24.36	23.27	24.16	25.42
	25.80	26.30	23.10	25.90	24.60	23.10	25.80	23.80	24.19	26.80
	26.32	26.23	22.62	26.77	25.58	23.35	25.79	24.86	25.00	27.10
A2	25.20	22.20	20.50	21.40	20.10	19.50	22.00	21.30	22.10	20.50
	22.40	23.20	20.90	22.40	21.40	21.30	22.30	21.50	20.40	22.00
	24.60	21.60	21.40	23.70	22.10	21.20	23.70	20.70	21.80	20.60
A3	20.40	19.70	20.90	22.40	23.00	19.40	22.60	17.40	19.60	21.60
	19.00	21.00	20.70	23.60	23.20	19.20	20.90	19.40	20.30	20.60
	21.10	20.30	21.90	23.40	22.20	18.40	21.40	18.70	19.10	22.90
A4	23.50	21.50	20.50	20.40	21.50	20.20	20.70	20.90	18.20	19.60
	23.40	22.60	21.40	21.10	21.80	20.90	21.40	20.70	19.50	20.90
	23.70	23.60	21.00	20.80	22.70	20.90	20.90	21.30	19.70	20.50
A5	21.53	20.95	18.98	21.91	16.47	17.78	19.53	19.85	18.00	19.96
	24.33	27.16	18.65	21.60	20.07	18.65	19.75	21.05	19.09	21.05
	23.40	22.36	19.64	23.01	20.84	19.42	20.62	20.62	20.35	21.82
A6	19.60	10.41	16.90	20.76	16.14	14.55	12.97	12.98	18.91	18.15
	19.55	17.14	17.17	19.03	17.15	15.55	15.47	15.47	17.33	14.60
	22.76	12.65	16.74	17.15	17.54	16.97	17.94	17.94	17.83	15.15
A7	18.66	16.64	16.94	18.55	14.90	14.52	18.55	16.46	17.95	16.64
	18.35	16.34	16.64	18.15	14.30	14.22	19.36	16.47	18.55	16.49
	19.16	18.00	16.13	17.75	14.60	13.28	17.65	16.69	18.09	16.74
A8	19.10	18.80	16.80	17.60	15.80	14.60	16.80	16.80	14.30	17.70
	19.60	18.00	17.30	17.10	16.20	16.70	16.60	16.50	15.90	18.50
	19.30	14.80	17.40	17.10	15.40	15.20	15.60	12.40	13.80	13.50
A9	21.60	13.70	17.90	19.70	17.40	17.30	20.50	17.40	19.60	16.50
	19.50	13.90	17.50	20.30	17.90	16.40	19.00	17.70	18.40	15.30
	22.00	19.00	17.30	19.90	17.20	16.10	19.80	16.40	19.80	17.80
A10	19.30	22.80	21.10	22.60	16.50	18.00	22.10	19.00	22.60	14.80
	21.10	22.70	22.00	23.40	22.10	21.70	22.70	22.00	23.20	16.40
	18.00	19.30	19.30	20.10	20.20	17.90	18.60	20.90	21.20	14.20
A11	22.10	19.87	18.71	20.88	19.30	10.70	19.20	11.85	19.48	16.80
	20.70	18.23	17.59	20.48	19.10	12.25	17.40	18.15	20.48	14.80
	20.40	20.09	17.77	19.96	19.90	15.00	17.35	11.55	20.25	15.70

图 1 数据分析

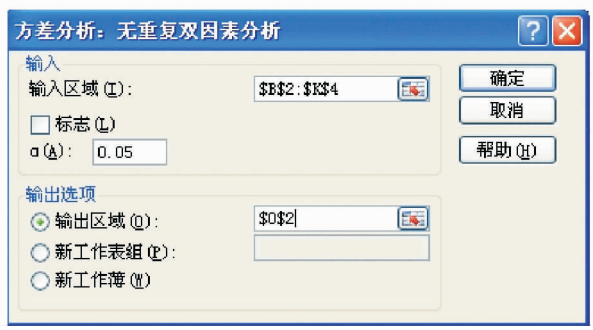

图 2 无重复双因素方差分析

表 3 A1 试点的方差分析

差异源	*SS*	*df*	*MS*	*F*	*P*-value
行	4.689 78	2	2.34 489	5.989 882	0.010 139
列	31.714 73	9	3.52 386	9.001 489	4.71E－05
误差	7.046 55	18	0.391 475		
总计	43.451 07	29			

出数据汇总表(表 4)。

2.3 品种效应与试点效应的计算 继续单击 Excel 菜单“数据”栏,选择 “数据分析”功能项, 再在对话框内选择“可重复双因素方差分析”分析项(图 3), 下一步单击确定按钮,出现下图的“可重复双因素方差分析”对话框(图 4);在输入区域,选中包括行 A1～A11、列 P1～P10 在内的所有原始数;由于每个试点区组为 3,因此样本行数填写 3;在输出区域指定任意单元格,单击“确定”可以得到 5 方差分析结果。

这种有重复双因素方差分析与一年多点试验设计略有

不同,后者仅在误差项进行再分解,即分解为“可解释的地点内区组间变异”和“随机误差”。这 2 个值通过表 4 已经得出,一年多点试验联合方差分析结果见表 6。

表 4　分析数据汇总

差异源	A1	A2	A3	A4	A5	A6	A7	A8	A9	A10	A11	Σ
区组内(行平方和)	4.690	2.184	0.294	3.749	18.794	6.523	0.148	17.589	4.485	39.561	0.080	98.096
误差平方和	7.047	15.689	12.586	2.925	26.223	67.605	5.149	24.278	24.909	27.639	47.170	261.220

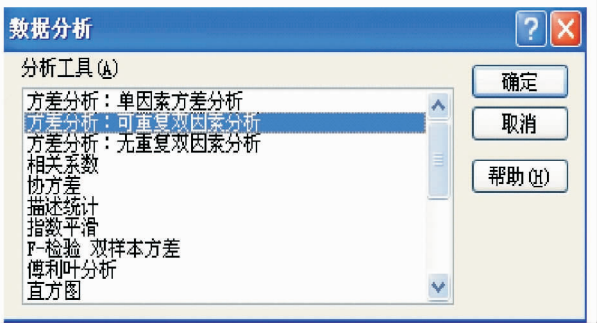

图 3　数据分析

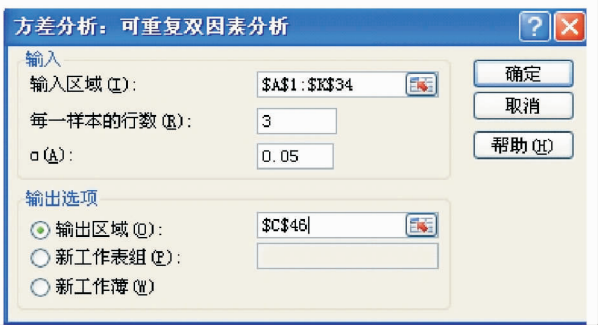

图 4　可重复双因素方差分析

表 5　有重复双因素方差分析

差异源	*SS*	*df*	*MS*	*F*	*P*-value	*F* crit
样本	2 058.89	10	205.80	126.00	3E－85	1.874
列	361.02	9	40.11	24.56	7E－29	1.923
交互	549.10	90	6.10	3.74	1E－15	1.326
内部	359.30	220	1.63			
总计	3327.30	329				

表 6　一年多点方差分析结果

变异来源	*df*	*SS*	*MS*
地点内区组	22	98.095 7	4.458 9
地点	10	2 057.890 0	205.788 9
品种	9	361.024 2	40.113 8
品种×地点	90	549.075 3	6.100 8
试验误差	198	261.220 0	1.319 3
总的	329	3 327.304 0	

3　讨论

依据模型不同,方差分析表中 F 有不同的计算方法。如果用固定模型,那么 F 值计算均以误差均方为分母;如果是随机模型,那么地点内区组、品种×地点的 F 值以误差均方为分母。品种、地点效应的 F 值以品种×地点为分母,但是随机模型中各个因素效应均为随机,实际上品种效应一般较固定,因此随机模型在分析中使用不太合理。如果选用固定模型,那么由于试点固定,所得结论仅限于试点应用范围有些狭窄,同时由于品种推广不可能只在这几个试点种植,所以固定模型在区试分析中具有一定的局限性。比较合理的模型是选用品种固定以及试点、区组随机的线性混合模型,这种模型的 F 值估算中地点内区组、品种×地点的效应是以试验误差为分母;品种效应的 F 值是以品种×地点互作为分母估算;地点效应的 F 值是以地点内区组为分母估算。Excel 经过简单操作后就可以输出方差分析的结果,但不能直接给出多重比较的结果。值得注意的是,多重比较要计算出各效应差数的标准方,计算公式 SE^2 =(品种×地点均方)/(ur),其中 ur 为品种样本容量[7]。在熟知其原理的基础上,利用 Excel 函数进行多重比较并不困难。虽然 Excel 统计部分的“加载宏”并非专为农业试验统计而写,没有关于田间设计的计算方法[8],但是将 Excel 软件提供的数据分析库中的几种方差分析工具组合起来,使用得当,同样可以用于农业上常用的随机区组试验的方差分析。

参考文献

[1] 谢德庆, 闫殿海. 作物区域试验中对品种评价方法的探讨[J]. 青海农林科技, 2006(2):4－7.
[2] 张群远, 孔繁玲. 作物品种区域试验的评价体系及方法[J]. 农业系统科学与综合研究, 2000,16(2):81－85.
[3] 马俊青,张静,周琳. EXCEL 在农林试验多因素统计分析中的应用[J]. 中国农学通报,2011,27(30):159－163.
[4] 马育华. 田间试验统计分析[M]. 北京:农业出版社,1985.
[5] 唐启义,冯明光. 实用统计分析及其 DPS 数据处理系统[M]. 北京:科学出版社,2002.
[6] 胡希远,李建平,宋喜芳. 空间统计分析在作物育种品系选择中的效果[J]. 作物学报, 2008, 34(3): 412.
[7] 潘家驹. 作物育种学总论[M]. 北京:农业出版社,1992:211－220.
[8] 霍世清,张静,冯岗. EXCEL 在裂区试验统计分析中的应用[J]. 中国农学通报,2011,27(30):159－163.