

利用 Excel 进行统计的方差分析

陈友春

(蚌埠医学院 计算机教研室, 安徽 蚌埠 233000)

〔摘要〕 目的:以方差分析为例,研究怎样用 Excel 完成简单统计分析. 方法:使用 Excel 的“数据分析”工具. 结果:单因素和双因素方差分析均可完成. 结论:利用 Excel 可以以较方便地完成简单统计分析.

〔关键词〕 因素;水平;方差分析

〔文章编号〕 1672-2027(2003)03-0046-04 〔中图分类号〕 TP391.13 〔文献标识码〕 A

在科学研究中,经常要对收集到的数据进行各种统计分析,由于运算量比较大,所以常使用一些统计分析软件. 如国际公认的三大著名统计软件:SAS、SPSS、BMDP^[1]. 但是一方面这些软件不容易获得;另一方面很多时候我们只是对数据进行简单的统计分析,如:t 检验、方差分析、相关分析及回归分析等. 作这些统计分析时,大多可使用 Windows98 下的 Excel 2000 来完成.

启动 Excel 2000,单击“工具”菜单中的“数据分析”命令可以浏览已有的分析工具. 如果在“工具”菜单上没有“数据分析”命令,应在“工具”菜单上运行“加载宏”命令,在“加载宏”对话框中选择“分析工具库”. 第一次使用,需要在光驱中放入 Office2000 安装光盘.

在科学试验中我们常常要研究参加试验的各种条件的改变对试验结果的影响,从中选出最好的试验条件组合,以达到最佳试验质量. 试验结果也称试验指标,试验中变化的条件称为因素. 因素在试验中所取的每一个状态称为因素的一个水平,方差分析是研究试验的各个因素各个水平对试验结果影响大小的一种方法^[2],根据情况可选择,单因素方差分析,无重复双因素方差分析、可重复双因素方差分析.

1 材料和方法

1.1 单因素方差分析

在试验中,只考虑某一因素对试验的影响,往往选择这个因素的几个不同水平进行试验,这样的试验称为单因素试验. 为判断该因素对试验结果是否有影响,应采用单因素方差分析.

例如:在进行某化学合成反应时,为了解催化剂对收率是否有影响,分别用 5 种不同的催化剂独立地在相同条件下进行实验,每种催化剂试验 4 次,得收率如表 1:

表 1 单因素试验的数据

Table 1 The data of simple factor experiment

催化剂	收	率
1	0.86	0.89
2	0.8	0.83
3	0.83	0.9
4	0.76	0.81
5	0.96	0.93

在工作表中输入上面的数据,比如数据区为 A₁ 至 E₆.

分析时,在“工具”菜单中,单击“数据分析”命令. 在数据分析对话框中,选择单因素方差分析,拉出单因

素方差分析对话框,其中有如下输入项(其他分析工具对话框内容和用法与之相似):

输入区域:在此输入待分析数据区域的单元格引用.该引用必须由两个或两个以上按列或行组织的相邻数据区域组成.可单击输入框右面的按钮,回到电子表格上自数据开始的单元格向结尾的单元格拖动.此时数据区域(B₂:E₆)自动进入输入域中.然后单击输入域右面的按钮,回到原对话框.

分组方式:如果需要指出输入区域中的数据是按行还是按列排列,请单击“行”或“列”,本例选择“行”.

标志位于第一行/列:如果输入区域的第一行中包含标志项,请选中“标志位于第一行”复选框;如果输入区域的第一列中包含标志项,请选中“标志位于第一列”复选框;如果输入区域没有标志项,则该复选框不会被选中,Microsoft Excel 将在输出表中生成适宜的数据标志.本例不选中该复选框.

Alpha(α):在此输入计算 F 统计临界值的置信度.Alpha 置信度为 I 型错误发生概率的显著性水平(舍弃真实假设).系统默认值为 0.05.

输出区域:在此输入对输出表左上角单元格的引用.当输出表将覆盖已有的数据,或是输出表越过了工作表的边界时,Microsoft Excel 会自动确定输出区域的大小并显示信息.

新工作表组:单击此选项,可在当前工作簿中插入新工作表,并由新工作表的 A1 单元格开始粘贴计算结果.如果需要给新工作表命名,请在右侧的编辑框中键入名称.

新工作簿:单击此选项,可创建一新工作簿,并在新工作簿的新工作表中粘贴计算结果.

给出结果如下:

表 2 单因素方差分析结果
Table The results of oneway analysis of variance

差异源	SS	df	MS	F	P-value	F crit
组间	0.044 27	4	0.011 067	10.34 346	0.000 317	3.055 568
组内	0.016 05	15	0.001 07			
总计	0.060 32	19				

结论:因为 P-value <0.01,故催化剂对收率有高度(**)显著影响.

1.2 无重复双因素方差分析

在实际问题中,影响试验结果的因素往往不只一个,而是两个或更多.当影响试验的因素为 A、B 两个时,对因素 A 和 B 的每一水平组合只作一次试验,这便是无重复的试验设计,此时为判断因素 A 和 B 对试验结果的影响,应采用无重复双因素方差分析.

例如:为考查蒸馏水的 pH 和硫酸铜溶液浓度对血清蛋白和球蛋白化验结果的影响,蒸馏水的pH(A)取 4 个水平,硫酸铜溶液浓度(B)取 3 个水平.在不同水平组合(A_iB_j)下各做一次试验,其结果列于下表中,试判断 A 和 B 两因素对化验结果有否影响.

表 3 无重复双因素试验的数据
Table3 The data of two factors experiment without repetition

	B ₁	B ₂	B ₃
A ₁	3.5	2.3	2.0
A ₂	2.6	2.0	1.9
A ₃	2.0	1.5	1.2
A ₄	1.4	0.8	0.3

在工作表中输入上面的数据.

分析时,在“工具”菜单中,单击“数据分析”命令.在数据分析对话框中,选择方差分析:无重复双因素分析.其中有如下输入项:

输入区域:本例输入区域为(A₁:D₅).

标志:本例选中该复选框.

其他输入项同单因素方差分析,给出如下结果(表 4):

表4 无重复双因素方差分析结果

Table 4 The results of two-way analysis of variance without reptition

差异源	SS	df	MS	F	P-value	F crit
行	5.289 167	3	1.763 056	40.948 39	0.000 217	4.757 055
列	2.221 667	2	1.110 833	25.8	0.001 13	5.143 249
误差	0.258 333	6	0.043 056			
总计	7.769 167	11				

结论:因为两因素的 P-value 均 <0.01 ,故蒸馏水的 pH 和硫酸铜溶液浓度对化验结果均有高度(**)显著影响。

1.3 可重复双因素方差分析

在两个因素的试验中,不仅需要考察每一个因素单独对试验的影响,有时还要研究两个因素联合对试验的影响,即两个因素的交互作用,常记作 $A \times B$. 这时 A 与 B 的各种水平组合($A_i B_j$)不能只做一次试验,而应进行重复试验,此时为判断因素 A 和 B 以及它们的交互作用对试验结果的影响,应采用有重复双因素方差分析。

例如:为探讨某化学反应中催化剂和温度对收率的影响,选取了甲、乙、丙 3 种催化剂和 4 个温度,对所有可能的组合在相同条件下各重复 2 次试验,所得结果见表 5,试判断催化剂、温度及它们的交互作用对试验结果是否有影响。

表5 有重复双因素试验的数据

Table 5 The data of repetitive two factors experiment

	70 C	80 C	90 C	100 C
甲	61	64	65	69
甲	63	66	66	68
乙	63	66	67	68
乙	64	67	69	71
丙	65	67	69	72
丙	67	68	70	74

在工作表中输入上面的数据。

分析时,在“工具”菜单中,单击“数据分析”命令。在数据分析对话框中,选择方差分析:可重复双因素分析。其中有如下输入项:

输入区域:本例输入区域为($A_1:E_7$)。

每一样本的行数:在此输入包含在每个样本中的行数。每个样本必须包含同样的行数,因为每一行代表数据的一个副本,本例输入 2。

其他输入项同单因素方差分析,给出如下结果(表 6):

表6 有重复双因素方差分析结果

Table 6 The results of repetitive two-way analysis of variance

差异源	SS	df	MS	F	P-value	F crit
样本	56.583 33	2	28.291 67	19.4	0.000 174	3.885 29
列	132.125	3	44.041 67	30.2	7.12E-06	3.490 3
交互	4.75	6	0.791 667	0.542 857	0.766 517	2.996 117
内部	17.5	12	1.458 333			
总计	210.958 3	23				

结论:由于两因素的 P-value 均 <0.01 ,故认为催化剂和温度对收率有高度显著影响。但由于交互作用的 P-value >0.1 ,故不能认为催化剂和温度间交互作用对收率有显著影响。

2 讨论

Excel 2000 中没有提供多个水平试验时的方差分析,在此本文不作讨论,由以上实例可见,Excel 2000

提供了常用的一些统计工具,利用这些工具将会大大减轻统计分析的计算量,借助 Excel 2000 中的公式和函数,尤其是大量的统计函数,也将帮助我们更方便地完成统计分析进程。当然,复杂的统计分析仍要借助专业统计分析软件包来完成。

参考文献:

- [1] 董大钧. SAS 统计分析(第1版)[M]. 沈阳:辽宁科技出版社,1995
- [2] 刘定远. 医药数理统计方法(第3版)[M]. 北京:人民卫生出版社,1999

How to Complete Statistic Analysis of Variance by Excel 2000

Chen Youchun

(Department of Computer, Bengbu Medical College, Bengbu 233000, China)

[Abstract] Take the analysis of variance as example, this article introduces how to complete statistic analysis by Excel 2000. Means: using "the analysis of data" tool which provided by Excel 2000 to complete analysis of variance. Result: both One-way analysis of variance and Two-way analysis of variance can be completed. Conclusion: Excel 2000 can be used for simply statistic analysis conveniently.

[Key words] factor; level; analysis of variance

(上接第45页)

而且操作方便;采用模块化设计,各系统可以单独运行,也可以相互配合使用,实现了医院网络系统的整体设计、分布实施,减少医院的投资压力;在保障医院系统安全运行的基础上,同时考虑医院的发展需要,体现了实用性、可发展性的明显优势。全面解决千年虫问题。

3.4 结构体系先进、支持多平台运行

采用先进的体系结构和网络环境,具有文件服务器和服务器/客户机两种模式,能在 NETWARE/DOS、WINDOWS NT/WINDOWS、UNIX 等多种平台上运行,并支持 INTERNET/INTRANET,以满足不同医院用户的需要。

4 结束语

系统充分体现了“以病人为中心”的现代医院管理思想与计算机信息技术的完美结合,系统覆盖了医院医疗诊治各环节和医院管理各主要方面,最大限度地提高医院运行的效率和医院管理水平,给医院带来了显著的经济效益和社会效益。

参考文献:

- [1] 黄梯云. 管理信息系统[M]. 北京:高等教育出版社,1999
- [2] 潘菡清,于燕莉,王晓燕,等. 药品库管理辅助程序的设计[J]. 中国医院药学杂志,1997,17(5):230
- [3] 张永明,陈德佳. 我院制剂室微机管理系统介绍[J]. 中国医院药学杂志,1997,17(3):134

The Exemplification of Designing about HIS

—Development of Hospital Information System

Lv Xiaoyan Li Xiangsheng

(Center of Computer, Shanxi Medical University, Taiyuan 030001, China)

[Abstract] Developed a method of designing information system based on an exemplification of hospital information system, if will be useful for the development of complex information system.

[Key words] computer; hospital information system; client/server