

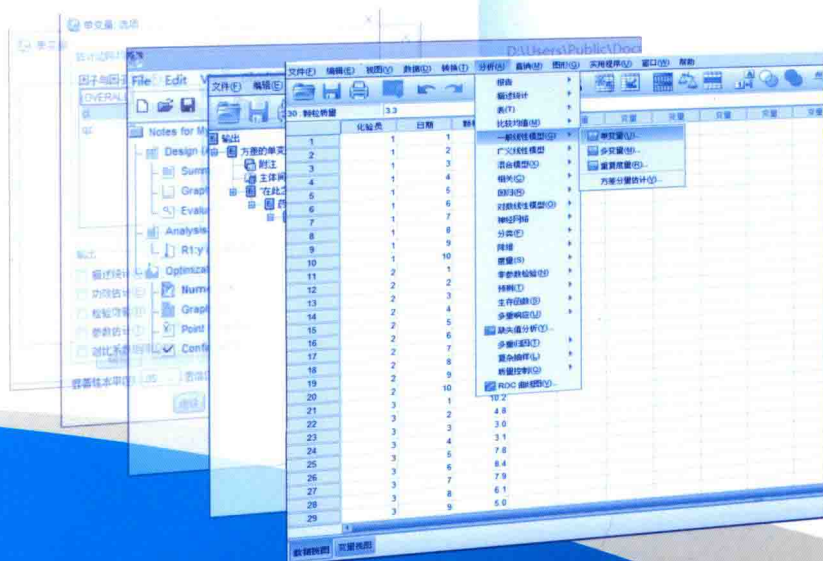
试验设计

与

软件应用

SHIYAN SHEJI YU
RUANJIAN YINGYONG

郭明 冯彬 管宇 主编



化学工业出版社

试验设计 与 软件应用

郭明 冯彬 管宇

主编



化学工业出版社

· 北京 ·

试验设计是一门以数理统计为基础的应用统计学分支学科,也是自然科学研究方法领域中重要的分支学科,正确的试验设计方案和对试验数据进行科学合理的统计分析是科学研究工作者必需具备的基本功。《试验设计与软件应用》分十二个章节,包括试验设计基础、方差分析、多元回归与相关分析、方差分析试验设计方法与统计分析、协方差分析、响应面试验设计与分析、混料试验设计与分析、均匀设计、聚类分析、规划、Plackett-Burman 试验设计与分析和 SigmaPlot 实例教程。

《试验设计与软件应用》可作为高等农林院校本科化学、食品、农学、生物专业及相关专业试验设计与统计分析课程的教材,也可供理、工、农、医等高等院校的相关专业作教材或教学参考书。

图书在版编目(CIP)数据

试验设计与软件应用/郭明,冯彬,管宇主编. —北京:
化学工业出版社, 2017. 7
ISBN 978-7-122-29984-0

I. ①试… II. ①郭… ②冯… ③管… III. ①试验
设计 IV. ①O212. 6

中国版本图书馆 CIP 数据核字 (2017) 第 139539 号

责任编辑: 李 琰 宋林青
责任校对: 边 涛

装帧设计: 关 飞

出版发行: 化学工业出版社 (北京市东城区青年湖南街 13 号 邮政编码 100011)

印 装: 三河市延风印装有限公司

787mm×1092mm 1/16 印张 13½ 字数 341 千字 2017 年 9 月北京第 1 版第 1 次印刷

购书咨询: 010-64518888 (传真: 010-64519686) 售后服务: 010-64518899

网 址: <http://www.cip.com.cn>

凡购买本书,如有缺损质量问题,本社销售中心负责调换。

定 价: 35.00 元

版权所有 违者必究

《试验设计与软件应用》编写组

主 编 郭 明

冯 彬

管 宇

编写人员 (按姓氏笔画排序)

王冰璇 卢闻君

李茜芸 姚献军

郭 明 冯 彬

管 宇

前 言

《试验设计与软件应用》是根据试验设计和统计分析的现状和趋势以及国内高校相关专业人才培养的实际情况进行编写的,在现有试验设计与统计分析教材的基础上,根据多年教学经验在介绍基本方法的基础上,突出试验设计方法和试验数据处理的实际应用,注重 Design-Expert、SPSS 电脑软件在试验设计和统计分析中的应用。

本书在试验设计基本理论、基本方法的基础上,突出 Design-Expert、SPSS 电脑软件在试验设计和数据处理中的实际应用,全面介绍了试验设计要解决的实际问题及其解决方法、原理及应用,并利用计算机软件进行实际的试验设计与统计分析。具体包括近年来试验研究中常用的、重要的试验设计和统计分析方法,如响应面试验设计与统计分析、均匀试验设计与统计分析、混料试验设计与统计分析、规划求解的方法等。理论与实践相结合,注重对学生创造性思维的培养和分析能力的提高。编写中以“精、全、新”为指导思想,在科学性、先进性、实用性上下功夫,力求概念准确、深入浅出、突出重点、语言简练,便于教学和阅读。与此同时,本教材着重培养学生的基本软件操作技巧、动手能力和思维能力。

本书具有显著的针对性和可操作性。因此,该教材的使用对于农林院校化学、应用化学、食品、农学、生物专业等相关专业教学过程中大学生创新思维的培养和创新能力的提高,具有极其重要的现实意义。《试验设计与软件应用》可作为高等农林院校本科化学、食品、农学、生物专业及相关专业试验设计与统计分析课程的教材,也可供理、工、农、医等高等院校的相关专业作教材或教学参考书。

限于作者水平,书中不足之处,敬请斧正。

编者

2017 年 4 月

目 录

第一章 试验设计基础	1
第一节 试验设计概述	1
一、试验设计的概念	1
二、试验设计的意义	1
三、试验设计常用术语	2
四、试验设计的程序与试验计划的制定	2
第二节 试验的基本要求	3
一、试验目的要明确	3
二、试验条件要有代表性	3
三、试验数据要有正确性	3
四、试验结果要有重演性	3
五、应当选择适当的试验指标,并有相应的数据分析方法	3
第三节 试验设计的基本原则	4
一、重复原则	4
二、随机化原则	4
三、局部控制原则——试验单位条件局部一致性	4
第四节 试验数据管理与准备	4
一、Excel 管理试验数据	4
二、SPSS 的数据格式	5
三、Excel 的数据管理应用	7
第二章 方差分析	10
第一节 方差分析的概念与基本原理	10
一、方差分析的概念	10
二、方差分析的基本原理	10
三、多重比较	13
第二节 单因素试验资料的方差分析与 SPSS 实现	16
一、各处理重复数相等的方差分析	17
二、各处理重复数不等的方差分析	19
第三节 两因素试验资料的方差分析与 SPSS 实现	21
交叉分组资料的方差分析	21
第四节 系统(组内)分组资料的方差分析与 SPSS 应用	37

一、次级样本含量相等的系统分组资料的方差分析	37
二、次级样本含量不等的系统分组资料的方差分析	42
第五节 方差分析的期望均方与方差组分估计及 SPSS 应用	46
一、概念	46
二、期望均方	46
三、方差组分的估计	49
第六节 多因素试验资料的方差分析与 SPSS 应用	53
第七节 数据转换	58
一、方差分析的基本假定	58
二、数据转换方法	58
第三章 多元回归与相关分析	60
第一节 多元线性回归分析	60
一、多元线性回归方程的建立	60
二、多元线性回归方程的假设检验	62
三、多元线性回归的估计区间	65
四、最优回归方程的选择	66
第二节 多元线性回归与 SPSS 应用	69
第三节 复相关分析	71
一、复相关系数的意义与计算	71
二、复相关系数的显著性检验	71
第四节 相关与偏相关分析	72
一、相关系数	72
二、偏相关系数的意义与计算	73
三、偏相关系数的显著性检验	74
第五节 相关分析与 SPSS 应用	76
第六节 通径分析与 SPSS 应用	78
一、通径系数与决定系数	78
二、通径系数的性质	80
三、通径系数的显著性检验	82
第七节 SPSS 通径分析	84
第八节 多项式回归与 SPSS 应用	86
一、一元多元多项式回归	87
二、SPSS 多元多项式回归	89
第四章 方差分析试验设计与统计分析	92
第一节 随机区组设计与 SPSS 应用	92
一、设计方法与要求	92
二、统计分析方法	92

第二节	拉丁方设计与 SPSS 应用	96
一、	设计方法	97
二、	统计分析方法	97
第三节	交叉设计与 SPSS 应用	100
一、	交叉设计概念	100
二、	设计方法 (单因素试验)	101
三、	统计分析方法	101
第四节	正交试验设计与 SPSS 应用	104
一、	正交试验设计的概念和特点	104
二、	正交表	104
三、	正交设计方法	105
四、	正交试验资料统计分析方法	105
第五章	协方差分析	126
第一节	协方差分析原理	126
第二节	协方差分析与 SPSS 应用	130
第三节	随机区组设计协方差分析与 SPSS 应用	133
第六章	响应面试验设计与分析	137
第一节	响应曲面分析法的基本概念	137
第二节	响应面模型	137
第三节	响应面试验设计与 Design-Expert 软件	137
第四节	响应面试验设计与分析实例	140
第五节	响应面方程应用	147
第七章	混料试验设计与分析	150
第一节	混料试验的回归模型	150
第二节	单纯形格子设计	151
第三节	单纯形重心设计	155
一、	单纯形重心设计试验方案的确定	155
二、	单纯形重心设计结果分析	157
第四节	Design Expert 的混料试验设计与统计分析	158
第八章	均匀设计	164
第一节	均匀设计概念与特点	164
一、	均匀设计概念	164
二、	均匀设计的特点	164

第二节 均匀设计方法	164
一、均匀设计表符号	164
二、均匀设计表结构	164
第三节 均匀设计试验数据的统计分析与 SPSS 应用	167
第九章 聚类分析	171
第一节 聚类分析的概念	171
第二节 聚类分析与 SPSS 应用	171
第十章 规划	176
第一节 规划的概念	176
第二节 Excel 规划求解	176
第十一章 Plackett-Burman 试验设计与分析	180
第一节 Plackett-Burman 试验设计与分析的概念	180
第二节 Design-Expert 的 Plackett-Burman 设计与分析	180
第十二章 SigmaPlot 实例教程	189
第一节 SigmaPlot 的功能	189
第二节 制作线图	189
第三节 制作三维散点图	195
第四节 二元系相图的绘制	199
附录	203
附录 1 r 与 R 的临界值表	203
附录 2 正交表	204
附录 3 均匀设计表	205
附录 4 字母对照表	206
参考文献	207

第一章 试验设计基础

第一节 试验设计概述

一、试验设计的概念

在生产和科学研究中,经常需要做试验,如何做试验,这里面大有学问,如果试验工作设计得好,试验次数不多,就能达到预期目的;试验工作设计得不好,会事倍功半,甚至劳而无功。设计一个试验要做很多工作,其中有两部分工作是非常重要的,一是试验方案的设计,二是试验结果的数据分析。为了更好地掌握试验设计的原理和方法,先必须了解试验设计的基本概念。所谓试验设计 (design of experiments),广义上是指试验研究的课题设计,也是整个试验计划的拟定,设计主要包括课题的名称、试验目的、研究依据、内容及达到的效果、试验方案、试验单位的选取、重复数的确定、试验单位的分组、试验的记录项目和要求、试验结果的分析方法、经济效益或社会效益估计、已具备的条件、需要购置的仪器设备、参加研究人员的分工、试验时间、地点、进度安排和经费预算、成果鉴定、学术论文撰写等内容;而狭义的试验设计是指试验单位(如果品储藏试验种类和品种)的选取、重复数目的确定、试验单位的分组和试验处理的安排。通常讲的试验设计指的是狭义的试验设计。

二、试验设计的意义

试验设计在试验研究中的意义主要体现在以下几方面。

(1) 确定试验因素对试验指标影响的大小顺序,找出主要因素。

(2) 提高试验研究的效度,明确试验因素之间相互影响的情况。试验的结果反映试验因素与试验指标间真实关系的程度称为试验效度。试验效度可以从以下两方面衡量:一是内在效度,指试验是否真的引起显著性差异,也就是要强调试验的重演性,内在效度高,重演性就好,它可以通过试验设计而得到显著提高;二是外在效度,指试验的结果能推广到什么范围,即强调试验的代表性问题,试验成果推广范围越广,其代表性就越强。所以在研制开发新产品的时候应具有与时俱进的思想,一个好的试验必须同时注意到内在效度和外在效度两个方面。

(3) 准确掌握最优方案并能预估或控制一定条件下的试验指标及其波动范围。

(4) 正确估计和有效控制、降低试验误差,从而提高试验的精度。

(5) 通过对试验结果的分析,可以明确进一步的研究方向。

合理的试验设计能避免系统误差,控制、降低试验误差,提高试验的精确性,保证试验的质量,从而对样本的总体作出可靠、正确的推断。

试验设计的任务是根据研究项目的需要,应用数理统计原理,作出周密安排,力求用较少的人力、物力和时间,最大限度地获得丰富而可靠的资料,通过分析得出正确的结论,明确回答研究项目所提出的问题。如果设计不合理,不仅达不到试验的目的,甚至可能导致整个试验失败。因此,能否合理地进行试验设计,已成为科研工作的关键。

三、试验设计常用术语

- (1) 试验指标 试验中具体测定的性状或观测的项目。
- (2) 试验因素 影响试验指标的原因。
- (3) 因素水平 试验因素所处的某种特定状态或数量等级。
- (4) 试验处理 事先设计好的、实施在试验单位上的具体项目。
- (5) 重复 在试验中,将一个试验处理实施在两个或两个以上的试验单位上。
- (6) 试验单位 在试验中能接受不同试验处理的独立的试验载体。

四、试验设计的程序与试验计划的制定

(一) 试验设计的程序

一般的试验程序可分为初级试验阶段、高级试验阶段和生产性试验阶段。它意味着由浅到深,由试验逐步到推广的几个阶段。

(1) 初级试验阶段 初级试验阶段是对某些品种或处理进行探索性试验,其特点是品种或处理数多,小区参试株数少,重复次数也少。主要有对照试验、比较试验、筛选试验等,其目的是在众多的因素中明确关键因素或优良水平。由于初级试验设计粗放、重复次数少,因此,试验误差较大,精度不高,只是探索其优劣而已。

(2) 高级试验阶段 高级试验阶段是经初级试验后筛选出来的品系或处理继续进行比较精密的一种试验,这一阶段主要是多因子的析因试验和优化试验,以深入分析主要因子的效应、交互作用以及寻找最佳的试验措施。

(3) 生产性试验阶段 生产性试验是从试验到大面积生产的过渡,是高级试验的继续和补充,是扩大试验规模,增加试验代表性的重要途径。通过生产上广泛的小规模试验,多点重复、反复验证,就能更有把握地推广试验结果,直接用于指导生产。

(二) 试验方案的制定与实施

(1) 明确试验目的与任务 在开始试验之前要提出所研究的问题,确定研究的对象和目标。

(2) 确定试验因子与水平 提出问题之后,接着要了解各种因子对试验结果的影响,分析因子的主次轻重,从中挑选出试验的关键因子与水平。

(3) 确定试验的总次数 要确定试验的总次数,先要明确试验中必须包括哪些水平组合,然后考虑试验所处环境因素的影响,再兼顾试验材料、人力、试验时间的长短等诸多因素。

(4) 挑选试验设计 要根据试验目的、试验条件、试验环境等因素来决定采用什么样的试验设计方法进行试验,同时要明确试验按怎样的顺序进行,采用怎样随机化的方法,在进行试验时要严格监控使试验计划的要求得到实现,并准确记录试验结果。

(5) 分析试验结果 在分析试验结果之前,有时需要对试验数据作适当的整理。统计分

析之后,要对分析的结果作出科学而又符合实际的解释,写出试验报告并提出建议。

第二节 试验的基本要求

一、试验目的要明确

明确选题,制定合理的试验方案,一是要抓住当时生产实践和科学试验中急需解决的问题,二是要考虑可能出现的问题。

二、试验条件要有代表性

试验条件应能代表将来准备推广试验结果的地区的自然条件、经济和社会条件。试验条件的代表性包括生物学和环境条件两个方面的代表性。生物学的代表性是指作为主要研究对象,如动物、作物品种、个体要有代表性,并要有足够的数量。例如,进行品种的比较试验时,所选样品的个体必须能够代表该品种,不要选择特殊性的个体,并根据个体均匀程度,在保证试验结果可靠性的前提下,确定适当的试验单位的数量。环境条件的代表性是指代表将来计划推广此项试验结果的地区的自然条件 and 生产条件,如气候、管理水平及设备。代表性决定了试验结果的可利用性,如果一个试验没有充分的代表性,再好的试验结果也不能推广和应用,就失去了实用价值。

三、试验数据要有正确性

试验结果的可靠程度主要用准确度和精确度进行描述。准确度是指观察值与真值的接近程度,由于真值是未知数,准确度不容易确定,故常设置对照处理,通过与对照相比来了解结果的相对准确程度。精确度是指试验中同一性状的重复观察值彼此之间接近的程度,即试验误差的大小,它是可以计算的。试验误差越小越精确。在进行试验的过程中,应严格执行各项试验要求,将非试验因素的干扰控制在最低水平,以避免系统误差,降低试验误差,提高试验数据的正确性。

四、试验结果要有重演性

重演性是指在相同条件下再次进行试验,应能够获得与原试验相类似的结果,即试验结果必须经受得起再试验的检验。试验的目的在于能在生产实践中推广试验结果,如果一个在试验中表现好的结果在实际生产中却表现不出来,那么,试验就失去了意义。由于试验中受试验单位之间的差异和复杂环境条件等因素影响,不同地区或不同时间进行的相同试验的结果往往不同;即使在相同条件下的试验,结果也有一定出入。因此,为了保证试验结果的重演性,必须认真选择供试单位,严格把握试验过程中的各个环节,在有条件的情况下,进行多年或多点试验,这样所获得的试验结果才具有较好的重演性。

五、应当选择适当的试验指标,并有相应的数据分析方法

分析试验结果最基本的统计方法是方差分析,因此试验结果数据必须满足方差分析的基本模型要求,如正态、独立、等方差等;若不能满足则需要采取相应的措施,如数据转换等。

第三节 试验设计的基本原则

一、重复原则

所谓重复 (replication) 就是指一个基本试验重复进行若干次, 即对应着某因子的诸水平或者某些因子的诸水平组合重复进行若干次试验。重复是科学调查结论的基本要求, 由于使用了重复这一手段, 在分析试验结果时, 就可以对试验误差做出估计。当因子诸水平或若干因子的诸水平组合的效应之间的差异超过误差时, 才能对它们的优劣作出比较和选择。此外, 重复能降低试验误差, 提高试验的精确度, 重复次数越多, 估计量的方差越小。但重复次数过多会带来试验时间长、经费支出高等诸多问题。

二、随机化原则

随机化 (randomization) 是指试验中每一个处理都有同等的机会实施安排在任何一个试验单元上, 即试验所使用的仪器、试验材料、试验操作人员以及试验单元等的执行顺序都要随机地确定。

随机化是试验设计所得数据使用统计方法进行分析的基石, 一方面随机化保证了试验结果是独立的随机变量, 在对效应作检验或估计时, 就可以应用数理统计学中独立样本的基本原理; 另一方面把试验进行适当的随机化亦有助于“平均值”可能出现的外来因素的效应, 有效地避免人为的主观性及外来因素对试验结果的影响, 以保证获得处理效应及误差变异的无偏估计。

三、局部控制原则——试验单位条件局部一致性

局部控制 (local control) 亦称区组化 (blocking), 是将试验单元按环境控制因子进行区组划分, 实行局部控制, 使同一区组内的试验单元间环境因子保持一致, 以保证同一区组中的局部范围内单元间误差的同质性。

在试验中, 当试验环境或试验单位差异较大时, 仅根据重复和随机化两个原则进行设计不能将试验环境或试验单位差异所引起的变异从试验误差中分离出来, 因而试验误差大, 试验的精确性与检验的灵敏度低。局部控制能排除试验材料间额外的、非处理引起的变异, 同时可减少试验误差, 增加处理效应估计的精度。实行区组化设计, 一方面要求同一区组内环境条件一致, 即同质性; 另一方面不同区组允许有异质的差异。因为单位组之间的差异可在方差分析时从试验误差中分离出来, 所以局部控制原则能较好地降低试验误差。

以上所述的重复、随机化、局部控制三个基本原则称为 Fisher 三原则, 是试验设计中必须遵循的原则, 其最终目的是为了提提高试验结果的精确度。只有正确地应用这三个原则, 并在试验中贯彻实施, 再采用相应的统计分析方法, 才能够最大限度地降低并无偏估计试验误差及无偏估计处理的效应, 从而对各处理间的比较得出可靠的结论。

第四节 试验数据管理与准备

一、Excel 管理试验数据

试验数据是说明事物本质的根本, 因此要经常整理试验所得数据, 有时要通过作表、作

图来说明问题。一般对试验数据的统计分析有规范化、标准化的要求，利用统计分析软件时，对数据格式也有要求，要按其要求准备好试验数据，进行统计分析。Excel 是微软公司开发的 Windows 环境下的电子表格系统，它是目前应用最广泛的表格处理软件之一，具有强有力的数据库管理功能、丰富的宏命令和函数、图表功能，直接用统计分析软件输入不是很方便，不如用 Excel 输入，因为 Excel 普及程度高，功能强，常规性的操作容易。

二、SPSS 的数据格式

【例 1-1】 比较施肥方法不同时，水稻平均产量是否一样，试验结果见表 1-1。

表 1-1 水稻 5 种施肥盆栽试验的产量结果

处理(施肥)方法	产量(x_{ij})			
1	24	30	28	26
2	27	24	21	26
3	31	28	25	30
4	32	33	33	28
5	21	22	16	21

表 1-1 已记录在 Excel 上，现简要介绍用 SPSS 给出统计结论的过程。

SPSS 要求数据格式是：处理号 数据（即每一个数据前标出它的处理号）运行 SPSS，见图 1-1。

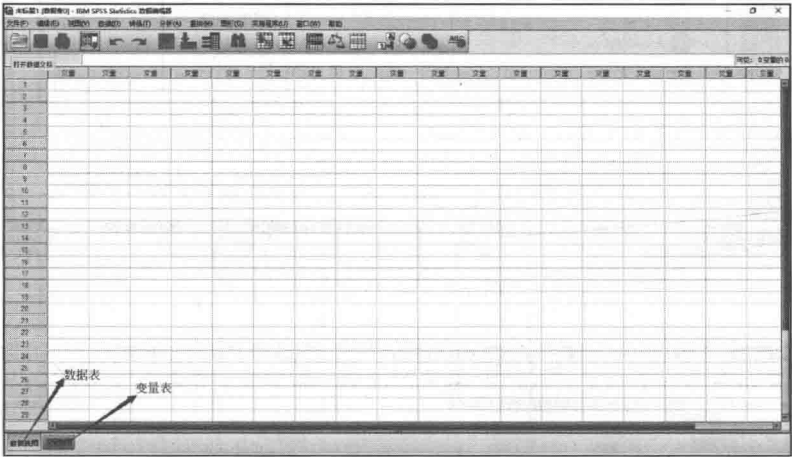


图 1-1 SPSS 数据表

图 1-2 是 SPSS 变量表，用于输入试验数据；单击变量（见图 1-3），用于给变量（或数据列）命名。例如，给第一个变量（或数据列）命名为 c_1 或处理，第二个变量（或数据列）命名为 x 或数据，然后单击变量表。

输入 c_1 、 x ，回到试验数据表（见图 1-4）。

输入试验数据表（见图 1-4），共 20 个（输入完毕后，即可运行分析程序给出结果）；或者利用 Excel 软件输入数据，输入完毕后，只要把他们复制到 SPSS 的 c_1 、 x 下即可。现利用 Excel 输入以上格式数据（见图 1-5 左端两列）。

将图 1-5 中数据复制到 A、B 两列下，再将 A、B 两列下的数据复制到 SPSS 数据表的 c_1 、 x 下，并进行统计分析。现简要介绍操作过程，首先运行 SPSS（见图 1-1），数据按要求输入完成后，运行方差分析，SPSS 统计分析结果见图 1-6。

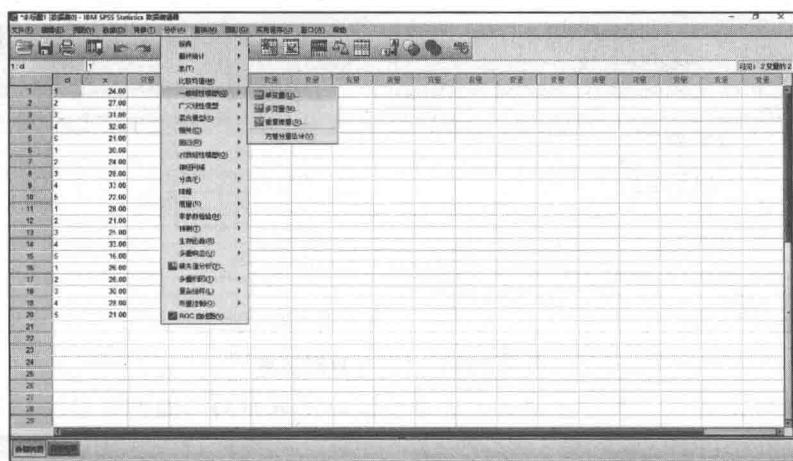


图 1-5 SPSS 数据格式和运行方差分析程序

主体间效应的检验

因变量: x

源	III 型平方和	df	均方	F	Sig.
校正模型	301.200 ^a	4	75.300	11.183	0.000
截距	13833.800	1	13833.800	2054.525	0.000
c ₁	301.200	4	75.300	11.183	0.000
误差	101.000	15	6.733		
总计	14236.000	20			
校正的总计	402.200	19			

a. $R^2 = 0.749$ (调整 $R^2 = 0.682$).

Duncan^{a,b}

C ₁	N	子集		
		1	2	3
5	4	20.0000		
2	4		24.5000	
1	4		27.0000	
3	4		28.5000	28.5000
4	4			31.5000
Sig.		1.000	0.055	0.123

图 1-6 SPSS 统计分析结果

由图 1-6 多重比较得出统计结论：肥料 4 与肥料 3 平均产量差异不显著，肥料 4 平均产量显著高于肥料 1、肥料 2、肥料 5 的平均产量，肥料 3、肥料 1、肥料 2 的平均产量差异不显著，而又都显著高于肥料 5 的平均产量。

三、Excel 的数据管理应用

(1) 试验数据见图 1-7 (a), 现要求将后几项数据补充完整, 操作过程如下: 先在 E3 中输入=B3*B3, 点击编辑栏前√号确认, …, 再在 G3 中输入=B3*C3, 点击编辑栏前√号确认, 见图 1-7 (b), 完成后见图 1-7 (c)。

(2) 求逆矩阵 操作方法见图 1-8 (a)、图 1-8 (b)。

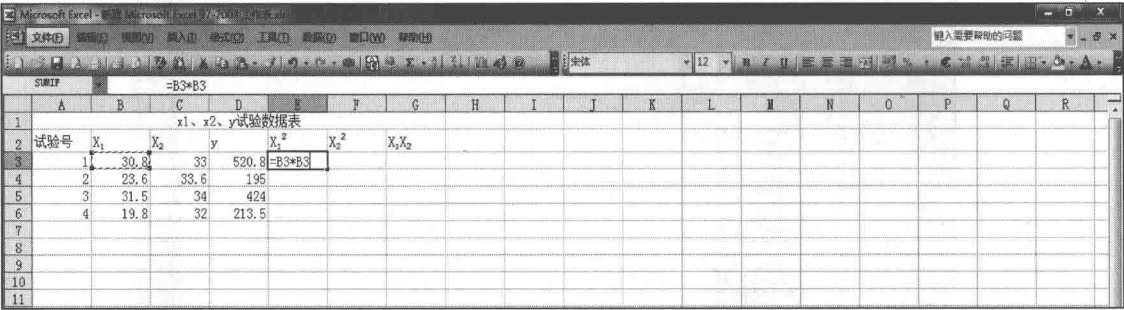


图 1-7 (a) Excel 数据运算

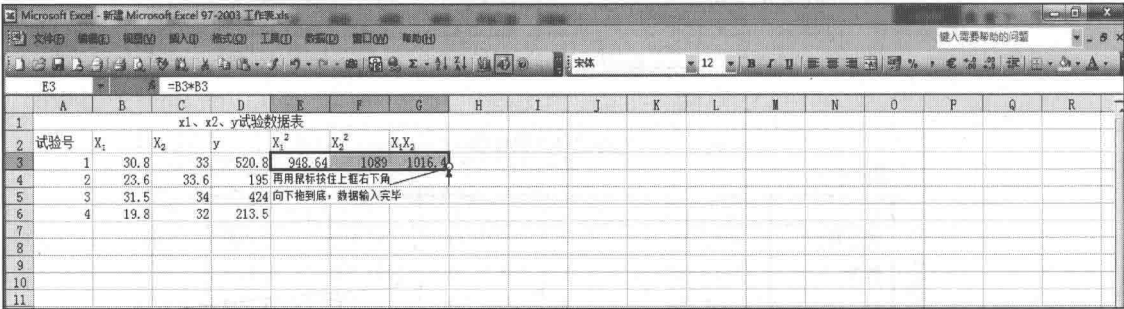


图 1-7 (b) Excel 相对地址使用

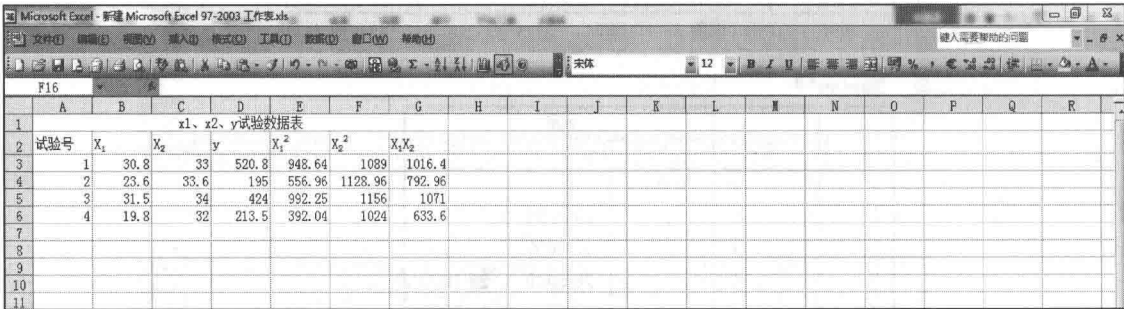


图 1-7 (c) Excel 数据复制

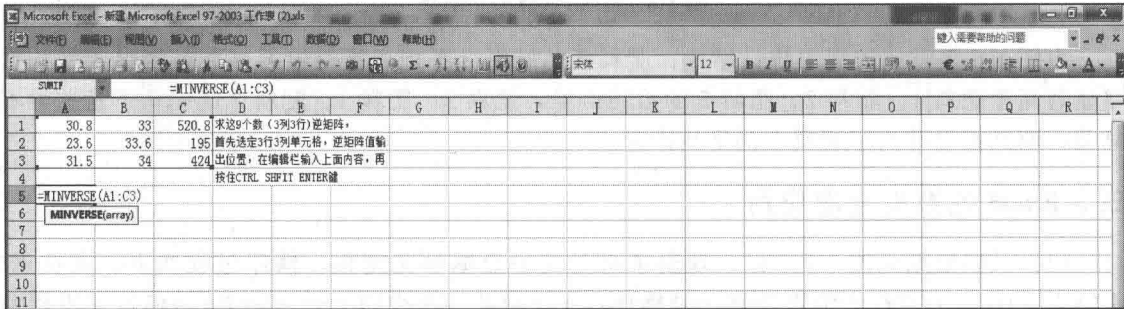


图 1-8 (a) Excel 上的逆矩阵操作