

试验设计与 Design-Expert、 SPSS 应用

徐向宏 何明珠 主编



科学出版社

www.sciencep.com

试验设计与 Design-Expert、SPSS 应用

徐向宏 何明珠 主编

目录 (CONTENTS)

第一章 绪论	1
第二章 试验设计的基本原理	10
第三章 正交试验设计	25
第四章 响应面法	45
第五章 试验设计软件的应用	65
第六章 试验设计在工程中的应用	85
第七章 试验设计在农业中的应用	105
第八章 试验设计在医学中的应用	125
第九章 试验设计在工业中的应用	145
第十章 试验设计在环境科学中的应用	165
第十一章 试验设计在材料科学中的应用	185
第十二章 试验设计在化学中的应用	205
第十三章 试验设计在生物科学中的应用	225
第十四章 试验设计在社会科学中的应用	245
第十五章 试验设计在心理学中的应用	265
第十六章 试验设计在教育学中的应用	285
第十七章 试验设计在政治学中的应用	305
第十八章 试验设计在历史学中的应用	325
第十九章 试验设计在地理学中的应用	345
第二十章 试验设计在环境科学中的应用	365
第二十一章 试验设计在材料科学中的应用	385
第二十二章 试验设计在化学中的应用	405
第二十三章 试验设计在生物科学中的应用	425
第二十四章 试验设计在社会科学中的应用	445
第二十五章 试验设计在心理学中的应用	465
第二十六章 试验设计在教育学中的应用	485
第二十七章 试验设计在政治学中的应用	505
第二十八章 试验设计在历史学中的应用	525
第二十九章 试验设计在地理学中的应用	545
第三十章 试验设计在环境科学中的应用	565
第三十一章 试验设计在材料科学中的应用	585
第三十二章 试验设计在化学中的应用	605
第三十三章 试验设计在生物科学中的应用	625
第三十四章 试验设计在社会科学中的应用	645
第三十五章 试验设计在心理学中的应用	665
第三十六章 试验设计在教育学中的应用	685
第三十七章 试验设计在政治学中的应用	705
第三十八章 试验设计在历史学中的应用	725
第三十九章 试验设计在地理学中的应用	745
第四十章 试验设计在环境科学中的应用	765
第四十一章 试验设计在材料科学中的应用	785
第四十二章 试验设计在化学中的应用	805
第四十三章 试验设计在生物科学中的应用	825
第四十四章 试验设计在社会科学中的应用	845
第四十五章 试验设计在心理学中的应用	865
第四十六章 试验设计在教育学中的应用	885
第四十七章 试验设计在政治学中的应用	905
第四十八章 试验设计在历史学中的应用	925
第四十九章 试验设计在地理学中的应用	945
第五十章 试验设计在环境科学中的应用	965
第五十一章 试验设计在材料科学中的应用	985
第五十二章 试验设计在化学中的应用	1005
第五十三章 试验设计在生物科学中的应用	1025
第五十四章 试验设计在社会科学中的应用	1045
第五十五章 试验设计在心理学中的应用	1065
第五十六章 试验设计在教育学中的应用	1085
第五十七章 试验设计在政治学中的应用	1105
第五十八章 试验设计在历史学中的应用	1125
第五十九章 试验设计在地理学中的应用	1145
第六十章 试验设计在环境科学中的应用	1165
第六十一章 试验设计在材料科学中的应用	1185
第六十二章 试验设计在化学中的应用	1205
第六十三章 试验设计在生物科学中的应用	1225
第六十四章 试验设计在社会科学中的应用	1245
第六十五章 试验设计在心理学中的应用	1265
第六十六章 试验设计在教育学中的应用	1285
第六十七章 试验设计在政治学中的应用	1305
第六十八章 试验设计在历史学中的应用	1325
第六十九章 试验设计在地理学中的应用	1345
第七十章 试验设计在环境科学中的应用	1365
第七十一章 试验设计在材料科学中的应用	1385
第七十二章 试验设计在化学中的应用	1405
第七十三章 试验设计在生物科学中的应用	1425
第七十四章 试验设计在社会科学中的应用	1445
第七十五章 试验设计在心理学中的应用	1465
第七十六章 试验设计在教育学中的应用	1485
第七十七章 试验设计在政治学中的应用	1505
第七十八章 试验设计在历史学中的应用	1525
第七十九章 试验设计在地理学中的应用	1545
第八十章 试验设计在环境科学中的应用	1565
第八十一章 试验设计在材料科学中的应用	1585
第八十二章 试验设计在化学中的应用	1605
第八十三章 试验设计在生物科学中的应用	1625
第八十四章 试验设计在社会科学中的应用	1645
第八十五章 试验设计在心理学中的应用	1665
第八十六章 试验设计在教育学中的应用	1685
第八十七章 试验设计在政治学中的应用	1705
第八十八章 试验设计在历史学中的应用	1725
第八十九章 试验设计在地理学中的应用	1745
第九十章 试验设计在环境科学中的应用	1765
第九十一章 试验设计在材料科学中的应用	1785
第九十二章 试验设计在化学中的应用	1805
第九十三章 试验设计在生物科学中的应用	1825
第九十四章 试验设计在社会科学中的应用	1845
第九十五章 试验设计在心理学中的应用	1865
第九十六章 试验设计在教育学中的应用	1885
第九十七章 试验设计在政治学中的应用	1905
第九十八章 试验设计在历史学中的应用	1925
第九十九章 试验设计在地理学中的应用	1945
第一百章 试验设计在环境科学中的应用	1965

科学出版社

北京

内 容 简 介

本书的特点是在介绍基本理论、基本方法的基础上,突出试验设计方法与试验数据处理的实际应用,如使用 Design-Expert 软件,利用它对相关问题进行试验设计,由其给出的试验设计进行试验,所得的试验数据输入软件内。Design-Expert 能够将试验数据自动进行处理,给出统计结果,包括 3D 图、极值或定值点等;SPSS 的使用使得试验数据处理很容易,为进行相关工作的人员提供了很好的试验设计和统计分析工具。

本书可为需要进行试验设计和统计分析方面工作的工作者提供参考。

图书在版编目(CIP)数据

试验设计与 Design-Expert、SPSS 应用/徐向宏,何明珠主编. —北京:科学出版社,2010.7

ISBN 978-7-03-027867-8

I. ①试… II. ①徐…②何… III. ①试验设计-应用软件-高等学校-教材
IV. ①O212.6-39

中国版本图书馆 CIP 数据核字(2010)第 105127 号

责任编辑:甄文全 席 慧 / 责任校对:邹慧卿
责任印制:张克忠 / 封面设计:耕者设计工作室

科学出版社出版

北京东黄城根北街 16 号

邮政编码:100717

<http://www.sciencep.com>

双青印刷厂印刷

科学出版社发行 各地新华书店经销

*

2010 年 7 月第 一 版 开本:787×1092 1/16

2010 年 7 月第一次印刷 印张:13 1/2

印数:1—3 000 字数:320 000

定价:27.00 元

(如有印装质量问题,我社负责调换)

编写委员会名单

主 编 徐向宏 何明珠

副主编 崔 岷

编 委(以姓氏笔画为序)

王 清 何明珠 陈德蓉

徐向宏 崔 岷 薛自学

前 言

试验设计是在数理统计的基础上逐渐发展起来的一门应用统计学分支学科,也是自然科学研究方法领域中比较成熟的分支学科。正确的试验设计是科学试验成功的必备条件,优良的试验设计方法及统计分析技术可保证试验结果推断的可靠性和科学性。正确设计试验方案并对试验所获数据进行科学的统计分析,是科学研究工作者必须具备的基本功。

本书是在甘肃农业大学为农科类及部分理科类专业编写的《试验设计与统计分析》讲义的基础上,考虑到现实要求,经必要的增删而完成的。本书的特点是在介绍基本方法的基础上,突出试验设计方法与试验数据处理的实际应用,如利用 Design-Expert 软件,对相关问题从试验设计到统计结果的输出自动完成,使得学生能够将时间用在实际问题的思考上,而不是如何设计、统计上。试验设计、统计分析是一种工具,而科学试验更需要多思考和实践。为此,本书涉及的问题是如何使用电脑软件进行试验设计和统计分析,希望在试验设计和统计分析的应用方面提供一种好的工具,更期望它能辅助科学试验。

本书得到了“甘肃农业大学硕士研究生重点课程建设项目”资助,编者在编写过程中参阅了许多专家的专著、论文、教材,在此向甘肃农业大学及各位专家表示谢意。

限于编者水平,错误、疏漏在所难免,敬请专家和读者指正,以便及时改正。

编 者

2010年3月14日

目 录

前言

第一章 试验设计基础	1
第一节 试验设计的概念.....	1
第二节 试验的基本要求.....	1
第三节 试验设计基本原则.....	2
第四节 试验数据管理与准备.....	3
第二章 方差分析	9
第一节 方差分析的概念与基本原理.....	9
第二节 单因素试验资料的方差分析与 SPSS 应用	14
第三节 两因素试验资料的方差分析与 SPSS 应用	19
第四节 系统(组内)分组资料的方差分析与 SPSS 应用	33
第五节 方差分析的期望均方与方差组分估计及 SPSS 应用	43
第六节 多因素试验资料的方差分析与 SPSS 应用	51
第七节 数据转换	56
第三章 多元回归与相关分析	58
第一节 多元线性回归分析	58
第二节 多元线性回归与 SPSS 应用	69
第三节 复相关分析	71
第四节 相关与偏相关分析	72
第五节 相关分析与 SPSS 应用	76
第六节 通径分析与 SPSS 应用	79
第七节 SPSS 通径分析	84
第八节 多项式回归与 SPSS 应用	88
第四章 方差分析试验设计方法与统计分析	96
第一节 随机单位组设计与 SPSS 应用	96
第二节 拉丁方设计与 SPSS 应用	101
第三节 交叉设计与 SPSS 应用	105
第四节 正交试验设计与 SPSS 应用	109
第五章 协方差分析	134
第一节 协方差分析原理.....	134
第二节 协方差分析与 SPSS 应用	138
第三节 随机区组设计协方差分析与 SPSS 应用	142

第六章 响应面试验设计与分析	146
第一节 响应面的概念.....	146
第二节 响应面模型.....	146
第三节 响应面试验设计与 Design-Expert 软件.....	147
第四节 响应面试验设计与分析实例.....	149
第五节 响应面方程应用.....	157
第七章 混料试验设计与分析	161
第一节 混料设计模型.....	161
第二节 单纯形格子设计与统计分析.....	162
第三节 单纯形重心设计与统计分析.....	166
第四节 Design Expert 的混料试验设计与统计分析.....	169
第八章 均匀设计	177
第一节 均匀设计概念与特点.....	177
第二节 均匀设计方法.....	177
第三节 均匀设计试验数据的统计分析与 SPSS 应用.....	179
第九章 聚类分析	184
第一节 聚类分析的概念.....	184
第二节 聚类分析与 SPSS 应用.....	184
第十章 规划	188
第一节 规划的概念.....	188
第二节 Excel 规划求解.....	188
第十一章 Plackett-Burman 试验设计与分析	194
第一节 Plackett-Burman 试验设计的概念.....	194
第二节 Design-Expert 的 Plackett-Burman 设计与分析.....	194
附表	204
附表 1 r 与 R 的临界值表.....	204
附表 2 正交表.....	206
附表 3 均匀设计表.....	207
主要参考文献	208

第一章 试验设计基础

第一节 试验设计的概念

一、试验设计的基本概念

生物统计是用部分来说明全体的,即通过样本了解总体。如何使选取的部分能够客观反映实际是统计科学领域研究的问题,这就是试验设计。试验设计是数理统计的一个重要分支。

二、试验设计的作用

试验设计的目的与作用就是避免系统误差,控制、降低试验误差,从而对样本所在总体作出可靠、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) 水平。试验因素所处的某种特定状态或数量等级。
- (4) 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。

第二节 试验的基本要求

一、试验条件要有代表性

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,所选择的个体必须能够代表该品种,不要选择性状特殊的个体,并根据个体均匀程度,在保证试验结果具有一定可靠性的条件下,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和生产条

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。

二、试验数据要有正确性

试验数据的正确性包括试验数据的准确性和试验数据的精确性。试验的准确性是指观测值与真值的接近程度,越是接近,准确性越高。试验数据的精确性是指试验数据相互的接近程度,越是接近,精确性越高。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。

三、试验结果要有重演性

重演性是指在相同条件下,重复进行同一试验,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。由于试验受试验单位之间差异和复杂的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。

第三节 试验设计基本原则

一、重复原则

重复是指试验中同一处理实施在两个或两个以上的试验单位上。设置重复的主要作用在于估计试验误差和降低试验误差。如果同一处理只实施在一个试验单位上,那么只能得到一个观测值,则无从看出差异,因而无法估计试验误差的大小。只有当同一处理实施在两个或两个以上的试验单位上,获得两个或两个以上的观测值时,才能估计出试验误差。同时,重复次数多可以降低试验误差。重复数的多少可根据试验的要求和条件而定。如果试验单位个体间差异较大,重复数应多些;差异较小,重复数可少些。

二、随机化原则

随机化是指在对试验单位进行分组时必须使用随机的方法,使试验单位进入各试验组的机会相等,以避免试验单位分组时受试验人员主观倾向的影响。这是在试验中排除非试验因素干扰的重要手段,目的是为了获得无偏的误差估计量。

三、局部控制原则——试验单位条件的局部一致性

局部控制是指在试验时采取一定的技术措施或方法来控制或降低非试验因素对试

验结果的影响。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。为解决这一问题,在试验环境或试验单位差异大的情况下,根据局部控制的原则,可将整个试验环境或试验单位分成若干个小环境或小组,在小环境或小组内使非处理因素尽量一致。每个比较一致的小环境或小组,称为单位组(或区组),即组内无差异,组间可能有较大差异。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、局部控制三个基本原则称为费雪(R. A. Fisher)三原则,是试验设计中必须遵循的原则,再采用相应的统计分析方法,就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试验数据,有时要作表、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。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,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

(单位:g)

处理(施肥)	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,现简要介绍使用 SPSS 给出统计结论的过程。

SPSS 要求数据格式是:处理号 数据(每一个数据前标出它的处理号,如 $\begin{matrix} 1 & 24 \\ 1 & 30 \end{matrix}$)运行 SPSS,见图 1-1。

图 1-2 是 SPSS 数据表,用于输入试验数据;单击变量表(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 cl 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。



图 1-1 运行 SPSS 图

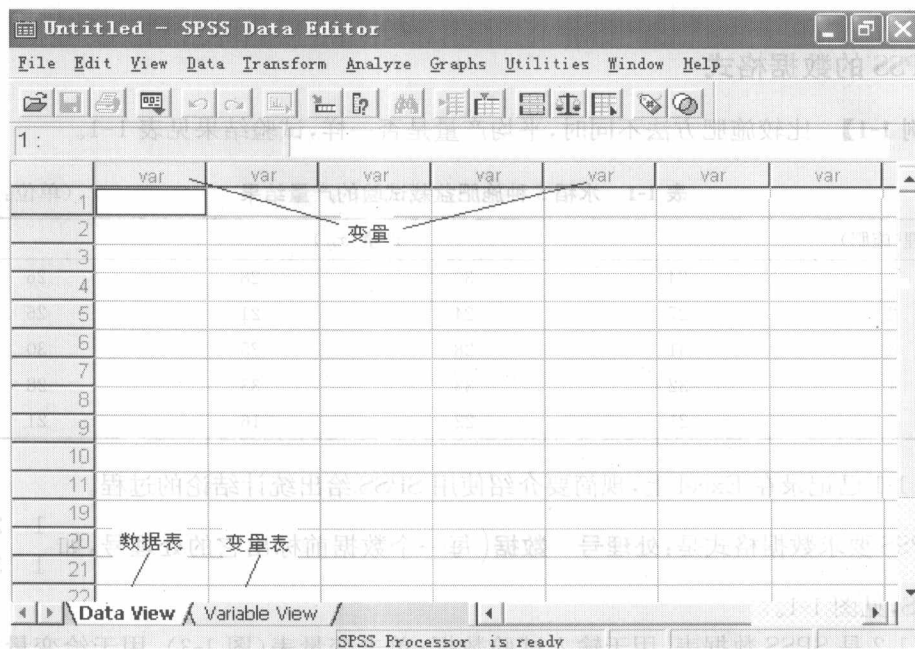


图 1-2 SPSS 数据表

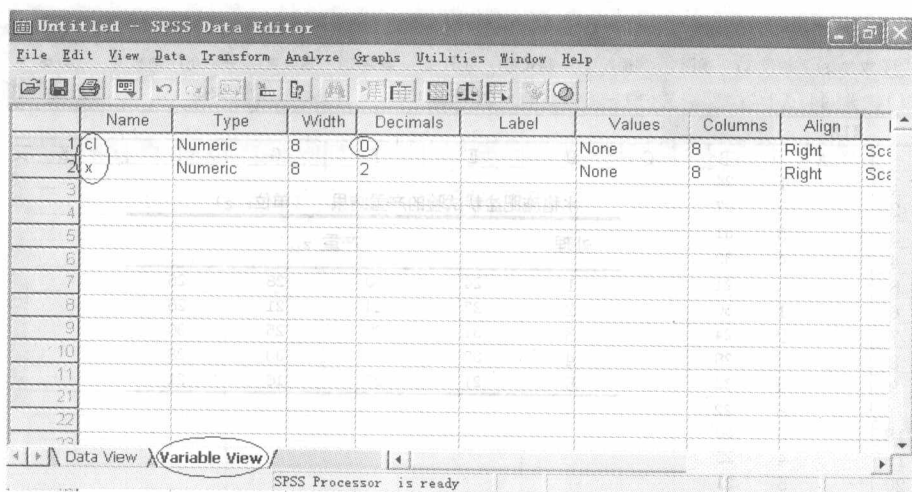


图 1-3 SPSS 变量表

输入 cl, x, 回到试验数据表(图 1-4)。

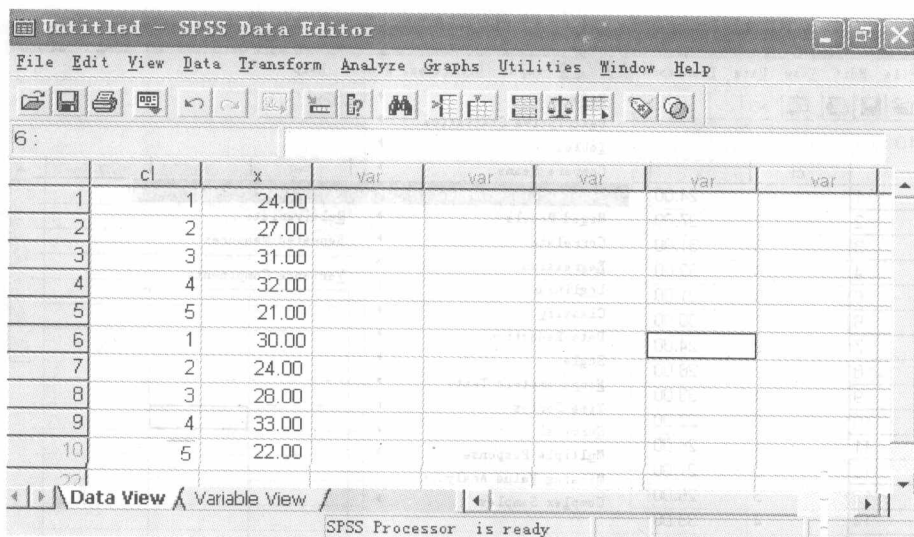


图 1-4 SPSS 数据表

输入试验数据(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,只要把它们复制到 SPSS 的 cl, x 下即可(因为 Excel 软件上已输入了试验数据)。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 cl, x 下,并进行统计分析。现简要介绍操作过程,具体操作方法见后面各章。首先运行 SPSS(图 1-1),数据按要求输入完成后,运行方差分析图 1-6(a)。

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

图 1-5 Excel 数据表图

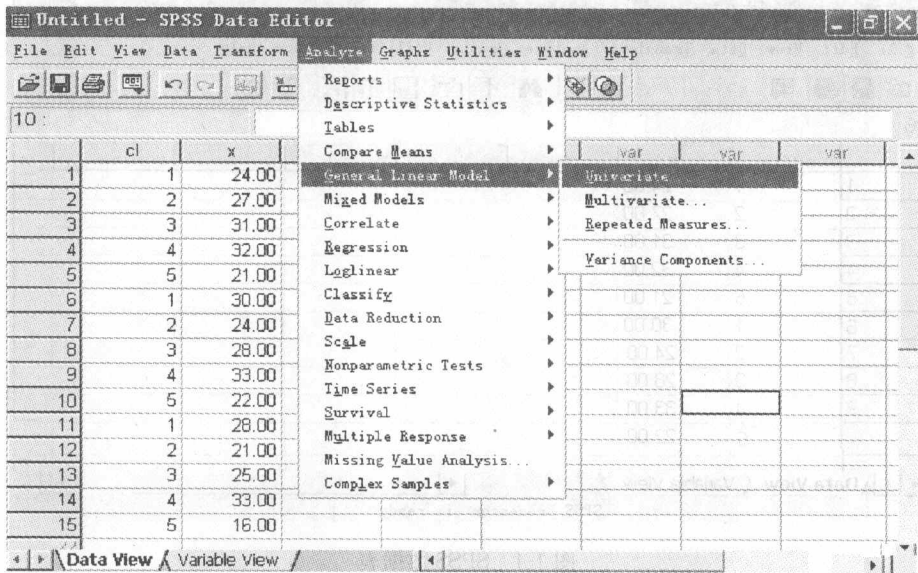


图 1-6(a) SPSS 试验数据格式和运行方差分析程序

SPSS 统计结果见图 1-6(b)。

由图 1-6(b)多重比较表得出统计结论: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 平均产量,肥料 3、肥料 1、肥料 2 平均产量差异不显著,而又显著都高于肥料 5 的平均产量。

Output - SPSS Viewer

File Edit View Data Transform Insert Format Analyze Graphs Utilities Window Help

Tests of Between-Subjects Effects

Dependent Variable: x

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	301.200 ^a	4	75.300	11.183	.000
Intercept	13833.800	1	13833.800	2054.525	.000
cl	301.200	4	75.300	11.183	.000
Error	101.000	15	6.733		
Total	14236.000	20			
Corrected Total	402.200	19			

a. R Squared = .749 (Adjusted R Squared = .682)

Duncan^{a,b}

cl	N	Subset		
		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	.055	.123

图 1-6(b) SPSS 统计分析结果

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), 完成后结果见图 1-7(c)。

Microsoft Excel - 图表.xls

文件(F) 编辑(E) 视图(V) 插入(I) 格式(O) 工具(T) 数据(D) 窗口(W) 帮助(H)

IF X ✓ / ✗ =B3*B3

	A	B	C	D	E	F	G
1	x ₁ 、x ₂ 、y 试验数据表						
2	试验号	x ₁	x ₂	y	x ₁ ²	x ₂ ²	x ₁ x ₂
3	1	30.8	33	520.8	=B3*B3		
4	2	23.6	33.6	195			
5	3	31.5	34	424			
6	4	19.8	32	213.5			
7							

Sheet1 / Sheet2 / Sheet3 /

图 1-7(a) Excel 数据运算

Microsoft Excel - 图表.xls

文件(F) 编辑(E) 视图(V) 插入(I) 格式(O) 工具(T) 数据(D) 窗口(W) 帮助(H)

IF X ✓ / ✗ =B3*B3

	A	B	C	D	E	F	G
1	x ₁ 、x ₂ 、y 试验数据表						
2	试验号	x ₁	x ₂	y	x ₁ ²	x ₂ ²	x ₁ x ₂
3	1	30.8	33	520.8	948.64	1089	1016.4
4	2	23.6	33.6	195			
5	3	31.5	34	424			
6	4	19.8	32	213.5			
7							

Sheet1 / Sheet2 / Sheet3 /

再用鼠标按住上框右下角，向下拖到底，数据输入完毕。

图 1-7(b) Excel 相对地址使用

试验号	x_1	x_2	y	x_1^2	x_2^2	x_1x_2
1	30.8	33	520.8	948.64	1089	1016.4
2	23.6	33.6	195	556.96	1128.96	792.96
3	31.5	34	424	992.25	1156	1071
4	19.8	32	213.5	392.04	1024	633.6

图 1-7(c) Excel 数据复制

(2) 求逆矩阵。操作方法见图 1-8(a)、图 1-8(b)。

求这9个数(3行3列)的逆矩阵。
首先选定3行3列单元格,逆矩阵值输出位置,在编辑栏输入上面内容,再按CTRL SHIFT ENTER 键

图 1-8(a) Excel 上的逆矩阵操作

所求的逆矩阵

5	-0.29017	-0.14154	0.421507
6	0.147205	0.127475	-0.23944
7	0.009753	0.000293	-0.00976

图 1-8(b) Excel 上的逆矩阵结果

利用 Excel 的函数功能、权柄复制功能能够非常方便地管理、准备试验数据及进行试验数据的统计分析工作。

第二章 方差分析

第一节 方差分析的概念与基本原理

一、方差分析的概念

方差分析主要用来判断多个总体平均数是否一致。例如,有 1、2、3 三个品种小麦在当地种植平均亩^①产是否一致? 如果不都一样,哪个品种平均亩产高? 又如,三种灌水方法 1、2、3,三种施肥方法 1、2、3,有 11,12,13,⋯,33 共 9 种不同搭配,哪个搭配平均产量高? 总体是研究对象的全体,如 1 品种亩产的全体是 1 总体,⋯,3 品种亩产的全体是 3 总体,它们的平均亩产称为总体平均数,设为 μ_1, μ_2, μ_3 , 总体平均数是常数,但我们不知道它的大小,为了比较它们的大小,就要试验或称为抽样,用样本平均数去估计总体平均数得出结论。一般的,1,⋯, k 个总体,要比较它们的总体平均数 $\mu_1, \mu_2, \dots, \mu_k$ 是否一致,就要抽样,得到试验数据,进行分析做出判断,这就是方差分析要解决的基本问题(图 2-1)。

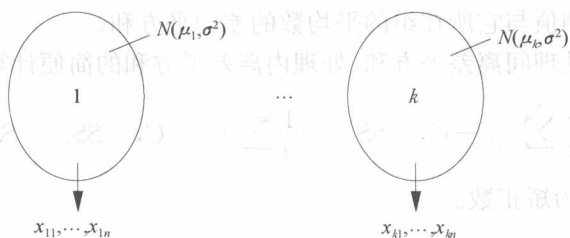


图 2-1 方差分析概念图

二、方差分析的基本原理

设有 k 个总体或处理,每个总体抽取 n 个个体,共有 kn 个观测值。试验数据如表 2-1 所示。

表 2-1 k 个处理试验数据表

处 理	观测值			合计 $x_{i.}$	平均 $\bar{x}_i.$
1	x_{11}	⋯	x_{1n}	$x_{1.}$	$\bar{x}_1.$
⋮	⋮		⋮	⋮	⋮
k	x_{k1}	⋯	x_{kn}	$x_{k.}$	$\bar{x}_k.$
合计				$x_{..}$	
平均					$\bar{x}..$

① 1 亩=666.7m²

表 2-1 中, $x_{i.} = \sum_{j=1}^n x_{ij}$ 表示第 i 个处理 n 个观测值的和; $x_{..} = \sum_{i=1}^k \sum_{j=1}^n x_{ij}$ 表示全部观测值的和; $\bar{x}_{i.} = \sum_{j=1}^n x_{ij} / n = x_{i.} / n$ 表示第 i 个处理的平均数; $\bar{x}_{..} = \sum_{i=1}^k \sum_{j=1}^n x_{ij} / (kn)$ 表示全部观测值的平均数。

1. 总离差平方和的剖分

总平方和为

$$\sum_{i=1}^k \sum_{j=1}^n (x_{ij} - \bar{x}_{..})^2 = \sum_{i=1}^k \sum_{j=1}^n (\bar{x}_{i.} - \bar{x}_{..})^2 + \sum_{i=1}^k \sum_{j=1}^n (x_{ij} - \bar{x}_{i.})^2 \quad (2-1)$$

设

$$SS_T = \sum_{i=1}^k \sum_{j=1}^n (x_{ij} - \bar{x}_{..})^2, \quad SS_t = \sum_{i=1}^k \sum_{j=1}^n (\bar{x}_{i.} - \bar{x}_{..})^2, \quad SS_e = \sum_{i=1}^k \sum_{j=1}^n (x_{ij} - \bar{x}_{i.})^2$$

则

$$SS_T = SS_t + SS_e \quad (2-2)$$

式中, SS_T 称为总的离差平方和, 是观测值与总平均数的差的平方和; SS_t 称为组(处理)间离差平方和, 是组的平均数与总平均数的差的平方和; SS_e 称为组内离差平方和或误差平方和, 是组内的观测值与它所在组的平均数的差的平方和。

总离差平方和、处理间离差平方和、处理内离差平方和的简便计算公式如下:

$$SS_T = \sum_{i=1}^k \sum_{j=1}^n x_{ij}^2 - C, \quad SS_t = \frac{1}{n} \sum_{i=1}^k x_{i.}^2 - C, \quad SS_e = SS_T - SS_t \quad (2-3)$$

式中, $C = x_{..}^2 / nk$, 称为矫正数。

2. 总自由度的剖分

在计算总平方和时, 资料中的各个观测值受 $\sum_{i=1}^k \sum_{j=1}^n (x_{ij} - \bar{x}_{..}) = 0$ 这一条件的约束, 故总自由度等于资料中观测值的总个数减 1, 即 $nk-1$ 。总自由度记为 df_T , 即 $df_T = nk-1$ 。

在计算处理间平方和时, 各处理均数 $\bar{x}_{i.}$ 受 $\sum_{i=1}^k (\bar{x}_{i.} - \bar{x}_{..}) = 0$ 这一条件的约束, 故处理间自由度为处理数减 1, 即 $k-1$ 。处理间自由度记为 df_t , 即 $df_t = k-1$ 。

在计算处理内平方和时, 受 k 个条件的约束, 即 $\sum_{j=1}^n (x_{ij} - \bar{x}_{i.}) = 0 (i = 1, 2, \dots, k)$ 。故处理内自由度为资料中观测值的总个数减 k , 即 $kn-k$ 。处理内自由度记为 df_e , 即 $df_e = kn-k = k(n-1)$ 。

因为

$$nk-1 = (k-1) + (nk-k) = (k-1) + k(n-1)$$

所以

$$df_T = df_t + df_e \quad (2-4)$$

综合以上各式得