

工业实验的设计 与分析

第二版

[英] O. L. 戴维斯 主编

0486441-0227256

化 工 学 社 出 版 社

本书系由著名学者O. L. Davies主编,作者通过来自生产实践的45个实例,对工业实验设计与分析原理和方法,作了深入找出的详细论述,不要求读者具有很多的数学知识。

本书共分十一章,其中主要内容包括平均数的比较、序贯检验、抽样和实验方法、随机化区组实验、不完全随机化区组实验、析因实验、最适条件的判定等。其中尤以四章的篇幅对各种类型的析因实验作了细致而有独创性的研讨。本书的另一特点是特别注意经济效益。

本书对从事工业技术、自然科学、企业管理等方面工作的工程技术人员,研究人员、管理人员以及有关院校的师生都有参考价值。

本书由杨纪珂同志校审

O. L. Davies
**The Design and Analysis of
Industrial Experiments**
Second Edition
Longman Group Limited 1973

工业实验的设计与分析

第二版

杨纪珂 刘垂甘 张璇 译

责任编辑: 雷延秀

封面设计: 许立

化学工业出版社出版

(北京和平门内大街七区十六号楼)

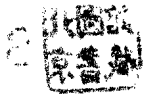
化学工业出版社印刷厂印刷

新华书店北京发行所发行

开本 $787 \times 1092^{1/16}$ 印张 $28^{1/4}$ 字数682千字印数1—5,670

1985年9月北京第1版1985年9月北京第1次印刷

统一书号15063·3680定价5.40元



译者的话

以数理统计方法为基础的实验设计与分析，起初是在生物科学上发展起来的，接着就迅速地转移到其它自然科学、技术科学、管理科学等的各个领域中去，并都已取得令人瞩目的成功。

在以要求高、批量大、连续性强因而必须充分考虑经济效益为特点的现代化工业生产中，应当怎样进行实验的设计与分析，本书是这方面不可多得的读物。

本书系由著名学者Owen L. Davies主编，并由包括他在内的六名专家共同撰写而成。作者们以帝国化学工业有限公司长期实践为依据，通过来自该公司的45个实例，对工业实验设计与分析的原理和方法，作了深入浅出的详细论述。实例虽来自化学工业，原理和方法却不失一般性。在严格保证原理和方法的科学性的同时避免深奥理论的空谈，充分重视经济效益而把试验与生产结合起来，灵活而巧妙的设计，细致而深入的分析，信息的充分利用，试验细节的清楚交待，等等，都提高了本书的价值。译者们深信，本书的翻译出版，将进一步推动我国工业实验设计与分析的工作而取得巨大的经济效益。

本书据朗曼公司专为帝国化学工业有限公司出版的1978年修订第二版译出。限于水平，在翻译中一定存在不少缺点和错误，欢迎读者批评指正。

中国科学技术大学 杨纪珂

安徽农学院 刘垂珩

西安重型机械研究所 张振

1982年3月1日

原作者序

帝国化学工业有限公司长期以来一直认为数理统计方法在工业研究与生产中占有重要的地位。为在它自己的机构内宣传并鼓励去使用这些方法, 1947年已安排出版过一本书^①。这本书不打算详细论述“实验设计”这样一个需要专著去详述的大课题。在这个领域中, 由于数理统计方法能以最少的试验从而也是以最小的花费去得出确切的结论, 这种方法已在工业研究中作出了重大贡献。因此, 本书将相当详细地对工业实验的设计与分析加以论述。

这一本书, 尽管自成体系, 却是上一本书^①的续篇。由于上一本书是以简单的方式陈述着手实验设计前所必须了解的基本数理统计方法, 故而上一本书是这一本书的一个很好的入门。这一本书中常需参照上一本书, 将在有关之处加以注明。

论及实验设计的著作大多基于农业研究所得经验, 但这些经验中的术语及实例则是其它领域(例如化学工业)研究工作者所不易理解的。然而, 更重要的是适于某一领域研究工作的实验设计并不一定适于另一领域的研究工作。既使某一用于农业研究的实验设计方法对于工业实验设计不可或缺, 非常重要的问题还在于过程的最适操作条件的索求。仅仅在最近才引进了合宜的设计, 并且本书的特色就是这些设计用于实际的方法及实例的充分论述。

象写上一本书那样, 作者们被允许随意地利用帝国化学工业有限公司所拥有的资料, 因此描述某一问题的每一个例子几乎都确系实践中所遇, 并给出了全部的工作细节。尽管实例来自化学工业, 但基本原理是用普通术语加以阐述的, 并可推广应用于其它工业领域。每个新的原理都用一个或更多的实例来说明。然而, 作者们要求自己只限于其价值已为自己的经验所证实的方法, 并且力求避免与所讨论问题无关的深奥理论。

本书原则上是为稍具数学数理统计学(mathematical statistics)知识的化学工业研究工作者而写, 同时希望本书能满足其它读者的需要。例如, 希望本书能为需要实际应用的大学和工学院的教师们以及打算进入工业界的学数理统计的学生们所适用。本书对那些打算在他们的特定领域内去对本书所给出的方法加以评价的管理人员将是有助的。

感谢为本书提供资料、研讨理论与应用以及评议初稿的帝国化学工业有限公司的有关人员; 感谢Biometrika托拉斯、G. A. Barnard教授、Research杂志主编、McGraw-Hill图书公司; 感谢本书出版者同意使用版权事宜。

* * * * *

节号、正文中的表号、公式号、图号及参考文献号都采用小数制度, 但不打算对不同的体系加以调配。例如表3.4是在§3.311中, 公式(3.1)则在§3.21中。节号用黑体字排出, 公式号则用斜体字, 其它数字就用正体字。

参考文献号放在方括号里面, [2.1]就表示“参阅Ronald A. Fisher的《实验设计》”。

本书末附有16种普通函数的表格, 表号为A、A.1、A.2、B、…、N。

* * * * *

自从本书1954年初版以来, 实验的统计设计已有许多重要发展。然而, 仍有必要特别对析因及析因实验作详尽的描述, 仍有必要对真实的工业实验作详尽的描述。因此, 1978年再次重印了第二版。内容大部未变。仅对早先读者们所指出的少数错误作了校正。所参考的《科研和生产中的统计方法》一书的章节, 这里指的是该书的第四版。

① 《科研和生产中的统计方法》, Owen L. Davies及Peter L. Golsmith主编, 朗曼公司出版。1947年初版。1976年修订第四版。

实例一览

第二章 简单比较实验的设计

橡胶的抗磨性能

对平均数增量的 t 检验

聚合物掺合器的效率

比率与标准值相比较

纱线的强度

对方差增量的 χ^2 检验

两种分析方法的可变性的比较

炸药起爆剂的检验方案

检验平均数差值的观测值的数目

橡胶的抗磨性能

检验平均数差值的观测值的数目

爆破用氢气机的效率

检验平均数差值的观测值的数目

极化炸胶的引爆

检验比率差值的观测值的数目

两种分析方法的精度的比较

所需观测值的数目

第三章 显著性的序贯检验

炸药起爆剂的序贯检验方案

平均数与标准值相比较, 标准差

化工厂中的改革

平均数的比较, 标准差已知—单侧备样假设

对合成纤维强度的研究

平均数的比较, 标准差已知—双侧备样假设

对纱线的染色性质的研究

平均数的比较, 标准差未知

来自两个过程的化工产品的质量

比率差值的比较

合成纤维纱线实验

缺陷频数的比较

对树脂膜强度的研究

方差缩减的检验

第四章 抽样和试验方法的研究

有机化学品的抽样和试验

抽样误差的确定, 试验误差及批间变差

白垩块密度的试验

实验室偏倚的确定, 抽样误差及试验误差

染料试验

实验室间偏倚的确定及试验误差

肥料的抽样及分析

方法间偏倚的确定, 抽样误差及试验误差

第五章 随机化区组及拉丁方

某种有机化学品的制造

在三个区组中五个处理的检验

纺织品的磨损试验

4×4 拉丁方

某种杀虫药粉的制备

7×7 拉丁方

铜管的退火

8×8 希腊拉丁方

合成纤维纱线物理性质中的变差

四个 3×3 希腊拉丁方的采用

第六章

轮胎的道路试验

在四个各容三个处理的区组中四个处理的不完全区组

橡胶的抗磨性能

在五个各容四个处理的区组中五个处理的尤登方

第七章 析因实验: 基本原理

某硝化过程的研究

2^3 析因设计

从一种异亚硝基乙基胺衍生物制备一种酞红衍生物

2^4 析因设计

第八章 各因素取两个以上水平的析因实验

硫化橡胶的抗磨性能

$5 \times 4 \times 3$ 析因实验

铝合金、抗腐蚀性能

$9 \times 4 \times 4$ 析因实验

一个 5×4 析因实验, 一个因素定性其它因素定量

青霉素、在种子罐生产中各种糖的不同用量的效应

6×4 析因实验, 一个因素定量

影响棉布染料的因素的检测

重复两次的 3^2 析因设计

第九章 析因设计中的混杂

某种有机化学品的产量

重复两次且混杂于四个区组中的 3^2 析因实验

各种因素对青霉素产量的效应

混杂于两区组中的 2^5 析因实验

各种因素对青霉素产量的效应

混杂于四区组中的 2^6 析因实验

第十章 分析因实验

染料制备时的过滤条件

四因素八观测值

某药品的产量

五因素八观测值

某基本染料的质量

五因素十六观测值

青霉素的产量

混杂于四区组中的五因素十六观测值

第十一章 最适条件的确定

最速上升法的例子

变动五个因素的某化学过程

第一次实验

第二次实验

接近平稳区域中产量曲面的探测

两个变量的例子

有三个变量的产量曲面的探测

原作者一覧

George E.P.Box,

Lewis R.Connor,

Wilfred R.Cousins,

Owen L.Davies,

Francis R.Himsworth,

George P.Sillitto.

目 录

实例一览

第一章 绪论	1
1.1 本书的目的	1
1.2 实验设计的性质和价值	1
1.3 观测数目的确定	4
1.4 序贯检验	4
1.5 抽样和检验方法的研究	5
1.6 随机化区组和拉丁方	5
1.7 析因设计	5
1.8 在实验设计中的序贯方法	6
第二章 简单比较实验的设计	7
2.1 引言	7
2.2 用一个简单例子说明设计原则	7
2.2.1 实验的介绍: 橡胶抗磨性实验	7
2.2.2 显著性检验的推理	8
2.2.3 随机化的必要性	9
2.2.4 采用成对试样的理由	10
2.2.5 实验误差合理估计值的选择	10
2.2.6 使实验配套的优点和缺点	11
2.2.7 平均差值的可靠性	12
2.3 其它简单比较实验的显著性检验和置信区间	12
2.3.1 单边和双边检验	12
2.3.2 比较实验的两种类型	13
2.3.3 平均数的比较, σ 已知 (正态曲线检验)	13
2.3.4 平均数的比较, σ 由试样估算 (t 检验)	14
2.3.5 比率或百分比的比较	15
2.3.6 方差的比较	15
2.4 控制第二类错误的风险	17
2.5 各种比较试验所需的观测数目	20
2.6 统计检验中的假定	24
参考文献	27
附录2A 误差为正态分布的假定	28
参考文献	32
第三章 显著性的序贯检验	33
3.1 引言	33

3.11 序贯检验	34
3.2 检验平均数的一个差异, 标准差已知	35
3.21 单侧备择假设平均值与某一标准值相比较	35
3.22 表达结果的另外一种方法	36
3.23 施行序贯检验的后果	37
3.24 单侧备择假设对于成对比较中平均数的一个差值的检验	39
3.25 双侧备择假设	41
3.3 当使用得自观测数据本身的标准差估计值时, 对平均数差值的检验	44
3.31 Barnard序贯 t 检验	44
3.32 Barnard序贯 t 检验. 双侧备择假设	47
3.4 比率差值的检验 (二项检验)	47
3.5 对频数差值的检验	50
3.6 对方差差异的检验	52
3.7 第三章讨论的序贯检验总结	54
参考文献	57
附录3A 简单假设的序贯检验	58
参考文献	59
附录3B 运算特性曲线或能力曲线	59
参考文献	60
附录3C 平均抽样数	60
第四章 抽样和试验方法的研究	62
4.1 引言	62
4.2 例4.1一种有机化学品的抽样和试验	66
4.3 方差估计量的精度	70
4.4 例子4.2检验白垩块密度	72
4.5 例子4.3检验染料	76
4.6 例子4.4一种化肥的抽样和分析	78
4.7 结论	82
参考文献	82
附录4A 各组中观测值数目不等时均方的期望	82
附录4B 由方差分析估计得出的两个方差之比的置信界限	83
附录4C 例子4.1的有关计算	84
附录4D 例子4.2的有关计算过程	86
附录4E 例子4.3的有关计算	89
附录4F 例子4.4的有关计算过程	91
第五章 随机化区组及拉丁方	93
随机化区组	93
5.1 引言	93
5.2 例子5.1一种有机化学品的生产	98
5.3 小结: 随机化区组	100

拉丁方	102
5.4 引言	102
5.5 例子5.2纺织品的磨损试验	105
5.6 正交方	109
5.7 3×3 拉丁方	112
5.8 拉丁立方	119
参考文献	120
附录5A 均方的期望	120
附录5B 例子5.1的计算过程	124
附录5C 例子5.2的计算过程	126
附录5D 例子5.3的计算过程	127
附录5E 例子5.4的计算过程	127
附录5F 例子5.5的计算过程	128
第六章 不完全随机化区组设计	130
6.1 引言	130
6.2 平衡化不完全区组设计的分析	132
6.3 对称平衡化不完全区组	137
6.4 尤登方	139
6.5 非平衡化不完全区组和格子方	143
6.6 格子方设计的两个重复	145
6.7 类拉丁方	147
6.8 类析因设计和类拉丁方设计的应用	147
参考文献	148
附录6A 平衡化不完全区组和尤登方一览表	148
附录6B 有缺失数据的双向表的一般处置	154
附录6C 非平衡化不完全区组	159
第七章 析因实验：基本原理	163
7.1 引言	163
7.2 因素	164
7.3 主效应和交互影响的解释	167
7.4 各因素仅取两个水平的的设计	169
7.5 一般 2^n 析因设计的分析	178
7.6 例子7.2 2^4 析因设计	183
7.7 有重复的 2^n 析因设计	186
参考文献	187
附录7A 析因实验中均方的期望	187
附录7B 2^n 设计的分析用Yates方法的系统验算	189
附录7C 一个及若干个均方的显著性	190
附录7D 交互影响均方合并成一个误差方差估计量	191
参考文献	193

第八章 各因素取两个以上水平的析因实验	194
8.1 引言	194
8.2 例子8.1一个三因素实验的简单示例	194
8.3 定量因素 多项式表示法	203
8.4 有两个定量因素的 $m \times n$ 析因设计	213
8.5 所有因素取三个水平的析因设计	218
8.6 3^3 析因设计	220
8.7 例子8.5 3^3 析因设计重复两次	220
8.8 3^4 设计和更高阶的设计	223
8.9 所有因素都是定量因素时,分析 3^k 析因设计的系统方法	223
参考文献	223
附录8A 定性因素析因设计的一般分析方法	223
附录8B 例子8.2的结果的详细分析过程	226
附录8C 用正交多项式计算定量因素时自由度的分解	228
附录8D 方差分析中平方和的除数	231
附录8E 例子8.4结果的详细分析	232
附录8F 3^3 析因设计的统计分析	234
附录8G 分析一项 3^k 设计的系统方法,其中所有因素均为定量因素,每一因素的水平均为等间距的	241
第九章 析因设计中的混杂在整个实验过程中无法保持一致条件时的析因实验工作 ...	245
9.1 引言	245
9.2 需要采用混杂原理的一些情况	251
9.3 在两个区组中混杂 2^k 设计的一般原理。在一项 2^k 析因设计里混杂某一给定交互影响的规则	253
9.4 2^k 析因设计高阶的混杂	257
9.5 3^k 析因设计中的混杂	263
9.6 3^3 设计中的混杂	265
9.7 取三个区组的 3^3 设计	268
9.8 3^3 设计中的混杂:九个区组各含三个观测值	269
9.9 3^4 析因实验中的混杂	271
参考文献	272
附录9A 例子9.1的统计分析详细过程	272
附录9B 2^k 析因设计中的部分混杂	274
附录9C 例子9.2的统计分析详细过程	276
附录9D 从有限群的观点引出混杂	277
附录9E 2^k 析因设计中的混杂系统	283
附录9F 一项 3^k 设计的例子:在三个区组中进行混杂及其详细分析过程	287
附录9G 3^k 混杂析因设计	290
第十章 分析因设计	294
10.1 引言	294

10.2 八个观测值的设计,各因素均取两个水平	298
10.3 析因设计与混杂之间的关系	301
10.4 一项分设计中效应的混杂。混淆	302
10.5 十六个观测值的设计	308
10.6 析因设计中的混杂	311
10.7 应用析因设计的讨论	314
10.8 3^n 析因设计中的分重复	318
参考文献	323
附录10A 十六个观测值时八个以下因素的分重复	324
附录10B 3^n 设计的分重复中的混淆	324
附录10C 析因设计结果的系统分析	328
第十一章 最速条件的确定	332
11.1 引言	332
11.2 寻找一个最大响应的方法	335
11.3 应用最速上升法找出接近平稳区域	339
11.4 一个最速上升的例子	342
11.5 接近平稳区域中产量曲面的探索。含两个变量的例子	347
11.6 多个变量的产量曲面的探测	355
11.7 探测三个变量的一个产量曲面的例子	361
11.8 一般结论	368
参考文献	368
附录11A 线性方程组的解 矩阵的逆 行列式的计算	369
参考文献	378
附录11B 响应曲面的拟合中最小二乘法的应用	378
参考文献	387
技术术语汇编	388
统计函数表	397
索引	421

第一章 绪 论

1.1 本书的目的

本书是论述实验工作的合理设计,是《科研和生产中的统计方法》(1947年版,1957年修订再版)一书的续编,后者主要研究如何用统计方法从现成数据中提取资料。本书不打算对工业调查研究做全面论述,只研究其中的一个方面,即为达成既定目的的一项复杂实验中各个项目的组织安排和结果的统计分析。本书不考虑施行实验的具体方法和科学研究的一般组织形式。在本书中,主要考虑与实验设计有关的分析并解释实验结果的统计方法的运用。

大部分化学和物理毕业生都不具备统计方法的知识,他们往往认识不到这些方法对其工作的重要性。对所介绍的统计方法,一开始只不过以为是惯常使用过的各种方法的变通而已——这种变通方法只有热心人才用,纯属可有可无。因此必须强调指出,凡是涉及到数据的问题,只要数据中包含有相当大的实验误差,则获得满意结果的唯一稳妥的处理办法便是统计方法,除此之外别无他择。如果一个实验者试图凭借自己的判断能力去评价他的工作结果,他所采用或应当采用的判据可能与那些统计检验所依据的判据相类似,可是不具备检验所需的数字条件。倘若所研究的效应大于随机变差,这种作法或许是满意的。但是当与效应相比,随机变差显得相当大时,这种直观判断就可能导致错误,用它来替代严谨的统计检验,就不合适了。在大学里,理工科大学学生们用高纯度的材料进行在严格控制下的高精度实验。但在工业生产过程中,这种理想条件往往是达不到的,即使能够达到,所取得的结果也往往不能直接应用于现实中所遇到的更为复杂的情况。因此,统计方法应当作为从事工业生产的科技人员所必须掌握的一门技术,用来有效地处理工业生产中的各种问题。

我们假定读者已熟悉了前……本书所述的较简单的统计技术,并将不时引证该书的相应章节。若想巩固对有关问题的记忆,请特别注意重温书中的第四、六、九章,这几章对本书使用的大多数统计检验作了充分的阐述。

对这门学科还不十分熟悉的读者,建议初读本书时暂先跳过第三、六章,因为这两章的内容不影响对后续章节的理解。相反,在第二章里包含着通贯全书的大量基本道理,应予仔细学习,并在需要时经常参阅之。

1.2 实验设计的性质和价值

化学工业中的实验工作五花八门,但其中很大一部分可归纳为下面几种类型的研究工作:

(a) **物理或物理化学研究** 这类实验通常涉及物质的基本常数和特征的精确测定。这种作业与大学中进行的研究工作类型非常接近,实验的设计一般取决于该作业的性质,除了某些涉及到误差评价的场合外,不常用到统计方法。

(b) **产品、原料等的常规分析** 化学分析中遇到的系统误差往往多于随机误差。这些系统误差常常随不同的分析者而异,或随同一分析者的不同分析时间而异。在这类情况下,即使由同一分析者重复几次该检验,也很难提高结果的精度,通常只能借以检出错误。除了实际化学分析误差外,在抽取样品和把一个样品划分为适于分析的若干小样的操作过程

中, 也会进一步引入相当可观的误差。所以, 查究抽样和分析的方案以了解所涉及的误差大小至关重要。对这类误差的研究需作小心的设计, 若想取得可靠的估计值, 其根本保证是在实验模型的选择和结果的解释中都采用统计方法。

(c) **材料试验** 对材料的机械和其它性质(如橡胶的抗拉强度)的试验中, 或对那些在定义上还不太明确的性质(如抗蚀性或耐久度)作评价时, 每个观测值往往都带有可观的随机误差。因此, 为了获得可靠的估计值, 必须从相当数量的观测值取其均值。凡是涉及这类试验的研究工作, 均需采用统计法则的合理设计。

(d) **化工过程的实验室研究** 几乎所有化工新过程和大部分旧过程的改革, 都是从小规模的实验室研究开始的。由于早期仅包括简单的和目的专一的实验, 所以对设计问题要求不高, 但迟早会达到必须就各种材料或条件的变化对过程的效应进行系统性研究的阶段。在这类系统性研究中, 统计学是科研工作选择最优设计的有力武器。

(e) **化工厂操作和生产过程的研究** 这类研究涉及过程效率的确定, 或在正常工作条件下过程操作的研究, 或由于各种操作条件或原材料的变化所起的效应, 等等。这类实验可能很费钱, 也可能需要许多人的协作, 因而必须认真计划以确保采用最合适的设计。

在研究任何指定的项目, 尤其是(d)型和(e)型的项目时, 研究人员可以或者从对所研究的特定过程有影响的各项基本关系入手, 或者采用某种更为经验性的办法, 直接在该过程中研究各种变化的效应。如果采用基本关系研究的方法, 研究人员可能对各种物理参量加以研究, 例如, 与该过程有关的各种化学反应的速率和平衡常数, 连同与之有关的各项传递系数, 渗透度等等; 并以这一切为基础, 试图推测出操作该过程的最佳条件和其中任一条件发生变化的效应。但是, 如果研究人员采用经验的方法, 他可以直接测定各种条件变化的效应, 而不去深究那些导致他所观测到的效应的确切机制。在工业研究中, 往往由于问题盘根错节非常复杂, 想要研究所观测的所有各种效应之潜在原因的工作量大得不可能完成, 因此就必须经常采用后一种经验方法。在这类经验法研究中, 以统计原理为根据的实验设计为取得所需的实践资料提供了经济的手段。即使所采用的是基本关系研究方法, 往往也会发现由于它是以实际系统的简化为根据, 所以要使实际过程最后调整到它的最佳条件, 仍然须依靠经验法完成之。

一项优良的实验设计, 能以最少的实验工作量取得所需的资料。为此需做三件事: 第一, 必须正确地列出要通过实验解答的各项问题; 第二, 必须在兼顾所要求的精度和可能碰到的实验难点的条件下正确地选用实验方法; 第三, 必须正确选定实验的一般模型, 即历次观测的数目、周期和相互关系。实验设计的统计理论所论述的就是这种包括在一组观测值中各项数据的数目及其相互关系的一般模型。利用数学理论, 可从而获得一个实验方案所提供资料的定量度量, 然后用它来与不同的方案进行对比, 以评价它们对任一给定项目的适用性。通过这类研究, 加上运用它们所取得的实践经验, 已经发展成为一门涉及到各种类型实验设计的学科, 用来指导科技人员从中选择出适合于特定项目的设计。

任何一组观测数据, 都是在所研究的各种条件下的一种模型或排列。如果我们在研究工作过程中, 以杂乱无章的方式取得结果, 则最终的排列可能也是不平衡和不规则的, 但若予以审查往往会发现当初要是设计得当, 用较少的工作量也能取得等量的资料。在这个领域里, 统计学的作用是提供各项判据, 使研究人员可据以判断所建议的某项设计的效率, 并为他们提供一些经过理论和实践证明了的、对处理某些类型的问题特别有效的某种标准设计模型。

用统计方法设计实验的一个附带的好处是, 迫使运用统计方法的研究人员必须预先考虑他们

在寻求那些目标,和必须采用那些步骤去探索这些目标。此外,还迫使他们考虑对于所有可能来源误差的大小。这种先期的考虑本身就具有重大价值,因为往往能够由此而引导研究人员认识到并从而规避可能的难点和失误,而且要不是这样做,这些难点或失误也可能要到研究工作的后期才被发现,那就晚了。

统计学的重要作用之一,是为确定观测数据的数目提供合理的依据。由于任何度量都带有一定程度的随机误差,因此往往必须综合好多观测数据的结果,才能使所得的资料具有足够的精度。在工业工作中,这种随机误差可能相当可观,因此在为一项指定目标而设计的研究工作中,要求以某些最小数目的观测值就可获得必需的精度。如果观测值超过了这个最小数目,那么原可更好地用之于其它工作的实验时间和精力、物力等于是浪费了。这相当于工程师们所谓的超需设计,在工业研究中,其虚耗的程度往往抵得上了一件工业设备的设计代价。反过来,如果实验属欠需设计的话,即观测数目太少,也可能得出虚假的结论。除非研究人员掌握了若干可供判断实验的正确数目用的明确判据,这类超需设计或欠需设计就属难免。经验证明,正确的设计类型和采取恰到好处处的观测数目两者的综合效果,可对一个科研项目的实验工作量带来经济上很大的节省。这种经济上节省的价值,可大大补偿为设计实验所花费的额外时间和脑力劳动而有余。由于任何研究工作的主要开支,几乎总是花在实验设计上,所以实验开支的减少会立即降低整个研究工作的成本。

为了对一项实验做出最佳最省的设计,需要具备有关的误差和需予侦察的效应大小的先期认识。乍看起来这是非常困难的,但实践中却极少要求一位研究人员去完成那种带有传奇性质的实验——有关误差的先期资料一无所有。这类资料有时是粗略的和不确定的,即使如此,利用这些资料进行设计,虽然可能比资料充分时的设计差一些,但无疑会比不加设计的实验好得多。如果实验工作的合理设计是有规律地定期进行的,应尽快建档收集这类资料,以供今后设计类似实验时参考。如果运用得当,合理设计的主要特点之一是鼓励研究人员在日常工作中充分运用全部知识和经验,辅助他规划好本职工作。

除去早期可资利用的资料外,伴随着工作的进展可以获得进一步的资料。例如,在化工研究中,各项试验或试制通常是一次一个或一次几个结果,因而在一项实验的过程中,先前试验的结果在其后的试验未开始前即可利用之。这时,我们可以运用已有的资料,确定实验设计和必要的进一步试验的次数。这种利用一系列观测中的先前部分获得的资料,设计其后的作业的概念,称作对一个问题的**序贯分析**。其最简单的形式是,先只做一次小规模中间实验,取得关于误差或效应的一般性质的概况资料,然后利用这些资料去设计整个实验。如果系统地运用序贯分析,其优越性更为明显。

在解释实验结果时,常常用到显著性统计检验。在这种检验中,通常假设某种想要探明的效应并不存在,然后去考查偏离此假设的差异是否有理由归因于偶然性的原因。如果发现出现这种差异的概率小于某一水准,例如小于 $1/20$,即可认为该假设不成立,并称该效应是显著的。但若概率大于这一规定的显著性水准,则称没有足够的证据去说明该效应的存在。这里当然必须应用各种显著性检验并审慎解释之。虽然本书中大多数设计方案原先都是为了使显著性检验套用得上而发明的,但是它们的价值并不靠这些检验,往往可省略检验而仍不降低这些设计的价值。当历次试验的变异性与所查究的效应相比显得更大(如用天然材料或生物体作实验所经常发生的情况)时,就需取得大量的观测数据,才能得出肯定的结论。这类情况下,在对任何效应确认其真实性之前,最好先做一下显著性的检验。此外,由于检验结果在很大程度上取决于随机误差的大小,因此最好采用这样的设计:可从实际的实验估计

出这些误差。

1.3 观测数目的确定

一般而论,科学实验的目的是以既定的精度去确定某项数量,或者去判定某项数量是否与一个标准值不同,或者与某个其它数量的差别超过一个规定值。在观测值伴有相当大的随机误差的情况下,可能就需要综合好些个历次分别观测的结果,以期获得精度较高的平均结果。倘若我们对误差一无所知,当然就不可能预计出实验所需的观测数目。不过,象经常所遇到的情况那样,只要知道了在各种有关度量值中所期望的误差大小,就不难使用基本的统计理论去预测出为确定一项具有既定精度的数量所需的观测数目。当我们希望检验的是某种差异是否存在时,问题就比较复杂了。利用前一本书中讲过的适当的显著性检验,可以保证将事实上并不存在的差异错判为存在的风险极小。但人们在施行这种类型的统计检验时往往忘记了他们冒有犯第二类错误的风险,即将一种事实上存在的,其大小足以产生重要实际意义的差异错判为不存在〔第二章〕。在工业研究中,避免错判差异不存在的重要性常常不亚于避免把并不真实存在的差异误认其有的重要性。例如,在一个化工过程中试验几种新的催化剂,并与常规使用的催化剂作比较,如果其中有一种新催化剂事实上并不比常规催化剂好,都被我们错判为一项革新,只需继续做些实验,这种错误一般就能很快校正过来,不会造成严重损失。但是,如果确有一种革新的催化剂被错判为不比常规催化剂好,就很可能不再对它作进一步的研究了,而该过程可能在很长时间里继续使用老的低效催化剂,由此带来的损失与上一种错误实际效果不大的革新所带来的较小损失相比,就相当不成比例。因此,在计划这类实验时,对这两种类型错误的风险均须给予应有的重视,实验的样品容量要足够地大,但又大得恰到好处,使两种风险都减少到可接受的水平。所冒的风险究竟允许多大,显然取决于两类错误的大概费用和进行一项较大规模实验的大概费用。倘若实验的费用很大,例如从实验的结果将决定是否兴建一个新车间的话,错判风险势必降至极低水平,但若实验结果将用于计划后继实验,则风险可取得大一些。

1.4 序贯检验

如果每做一次试验都要把它的结果检查之后再做下一次试验是不现实的或不方便的,而整个实验又必须一气呵成的话,则必须果用上述方法。但若观测是一个接一个或按小组进行的,则使用显著性序贯检验可能更经济。对一连串的结果,每到手一个就施行一次适当的检验,一俟掌握了足够的根据能够表明或是存在差异、或是所有可能存在的差异不大于某些给定值时,就立即停止实验〔第三章〕。在此类检验中,实验工作要持续到接受了原假设或供选择假设中的一个后方可停止。在理论上,对所需的观测数目并无确定的上限,但平均说来这个数目约为有固定规模的一项实验所需的一半左右,偶而也会遇到一些所需观测数目较多的实验项目。这种所需观测数目的不肯定性,并不象人们想象的那么严重,因为在实践中,做出判断需用的观测数目,比有固定规模的实验需用数目多出一倍的情况难得发生,而绝大多数序贯检验的迅速终结,又很好地补偿了这种偶遇的大量观测所带来的麻烦。当然,任何序贯实验也可按给定的观测数目而终结,不过如此行事会使得出错误结论的风险略有增加。在这样的情况下,判据可资用于得出一个结论。