

数理统计的原理和方法

适用于生物科学

[美] R. G. D. 斯蒂尔 著
J. H. 托里

科学出版社

数理统计的原理和方法

适用于生物学

〔美〕 R. G. D. 斯蒂尔 J. H. 托里 著

杨纪珂 孙长鸣 译

科学出版社

1979

内 容 简 介

本书是继 G. W. 斯奈迪格《数理统计方法》一书后的比较完善的一本有关数理统计原理和方法的著作。本书的特点是把有关的原理用深入浅出的笔法剖析得非常透彻,便于不是以数学为专业的读者学习与掌握。内容包括观测数据、概率、抽样方法、统计比较、实验设计、方差分析、回归、协方差分析、属性统计、列联表、离散型分布、非参量统计等,其中以六章的篇幅对数理统计方法中最重要的方差分析(包括协方差分析)作了详尽地介绍,有很大的实用价值。

书中所举例子多为生物学、农业及医学方面的,所以对从事生物学、农业、医学的研究人员及有关的大专院校师生更有参考价值。

Robert G. D. Steel and James H. Torrie

PRINCIPLES AND PROCEDURES OF STATISTICS

WITH SPECIAL REFERENCE TO THE
BIOLOGICAL SCIENCES

McGraw-Hill Book Co., Inc. 1960

数理统计的原理和方法

适用于生物科学

[美] R. G. D. 斯蒂尔 J. H. 托里 著

杨纪珂 孙长鸣 译

*

科学出版社出版

北京朝阳门内大街 137 号

上海商务印刷厂印刷

新华书店北京发行所发行 各地新华书店经售

*

1979 年 5 月 第 一 版 开本: 850×1168 1/32

1979 年 5 月第一次印刷 印张: 19

印数: 0001—24,300 字数: 491,000

统一书号: 13031·305

本社书号: 475·13—6

定 价: 2.30 元

译 者 的 话

本书是美国威斯康辛大学的教科书,系1960年出版,但至今国内外生物科学工作者仍在广泛地应用。因此,我们把它翻译出来,介绍给读者。

该书的主要特点是详细透彻地说明了数理统计的应用方法;深入浅出地阐述了数理统计的数学原理,特别是书中对在生物学各分支中所遇到的实际统计问题都举出了较好的实例,因而受到了生物科学研究工作者的欢迎。它既可作为高等院校的教科书和参考书,又是生物科学研究工作者不可缺少的工具书。

在翻译过程中,根据具体情况,作了必要的删改,错误之处也予以纠正。

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

中国科学院生物物理研究所 孙长鸣

1978年1月20日

序 言

这本教科书的讲义是作者二人在威斯康辛大学任教时编写的。经过师生的讨论，认为在为农学院研究生进行数理统计的教学工作中有三个应予注意的要点。

首先，必须在有限的时间之内顾及到数理统计原理与方法的各个方面。大多数的学生只能安排一两个学期的时间来学数理统计，因为除了那些以数理统计为主要或次要专业的学生外，数理统计一般是门辅助课程。

其次，应设法尽可能快地把方差分析法教给学生。一般研究生在开始时是缺乏数理统计训练的，但如兼任了助教，就得负起田间、温室、动物房或实验室的实验工作的责任，实验所得的度量数据需用方差分析法来计算。此外，研究生论文的问题也需用实验设计和方差分析的知识，而他可能还缺乏这方面的知识。

最后，需要用非数学的方法来教这门课程。应用数理统计的教师往往发现学生的数学水平较差，并且有点害怕数学。所以我们决定免掉代数上的推演。这样做是否已完全取得成功，需视学生对第二十章里我们介绍二项分布的方式的感受如何而定。

跟数学家一样，统计家也使用符号与添标，作为对常遇的并且是重要的情况的冗长描述的替代。似乎没有必要不去使用它们。求和的代号有两个： Σ 以及“黑点添标”。用希腊字母和英文字母来区别总体参量与样本统计量。希腊字母列出在附表中。

在数理统计的教学中，我们三个主要的目的：

第一，把大量的数理统计技术介绍给学生，这些技术在研究工作中都是用得着的。曾对学过数理统计的毕业生作过一次小小的调查，发现任何一种数理统计方法被采用的广泛程度取决于在教学中的注重程度。很少人在学过这门课程后再去自学其它方法。

第二, 对学生在如何进行科学实验这个问题上灌输一些成熟的概念. 应该把学生训练到这样一个程度, 使他能够认识到诸如 Salk 的小儿麻痹症疫苗的试验的一部分并不与试验的目的有关; 诸如一项实验不一定是为了确定相对精确度, 用来为不同的样本含量度量其平均数; 诸如图表、平均、比例、百分率以及非随机的抽样等会经常粗率地为人所误用; 等等.

第三, 对学生在如何分析实验数据这个问题上灌输一些成熟的概念. 例如, 一个学生应该有能力精审地去评估数字, 从自然变差的角度认识到它们的限度与缺点, 但也充分地认识到它们的功效.

我们用原先的讲义开始授课, 随着在教学经验上的增进, 尽可能地作了删节和修改. 我们对数学模型予以相当的重视. 又增加了一些来自原作者的统计方法, 例如在各处理平均数间的检验方法. 我们还加进去一些学生需知道的有参考价值的材料. 我们体会到没有一本教科书可能是完整无缺的, 所以我们最后又重写了一遍, 然后把稿子付印.

这本教科书的排列次序是按课程而排的. 在教学方法与课程内容上允许有相当大的周旋余地.

在头五章里, 介绍了数理统计的概念, 为理解以后各章中许多数理统计方法提供一个基础. 这些概念包括母分布与导出分布、概率、置信区间、假设的检验、I 型与 II 型误差、辨别能力、样本含量, 以及方差分析. 由于这些概念对学生来说, 通常是比较新鲜且难懂的, 所以, 我们反复地并不厌其详地把它说明白.

在介绍数理统计方法的各章里, 我们也企图指出其内在的逻辑性. 这对实验设计与线性回归而论, 部分是通过线性模型的讨论来进行的.

第六章的位置放得有些不自然. 此时学生当然还没有深入学习它的准备知识. 有些教师希望把它放在课程的后期去教, 有些教师在教方差分析之前先在此简介一下, 等到教过第十二章后回头再详教这一章.

在方差分析之后讨论了离散型的变量,并顾及了协方差。在讨论到 χ^2 在涉及离散型变量的问题的应用时,强调地指出了 χ^2 是个连续型分布的事实,从而指出这个也被称为 χ^2 的检验判据只是近似地按 χ^2 分布而分布的。对二项分布的讨论故意地被留下,直到讲过两章离散型数据(包括二项分布数据在内)之后再提。这样做有两个理由:首先,如果先讲二项分布,有可能过分强调了计数数据的二项分布性,损害了 χ^2 检验判据的非参量性的潜力。其次,也许有可能使多一些的学生,在与 χ^2 作为检验判据和计数数据打了一阵交道后,对第二十章仍感兴趣。

总之,这本教科书是为在数学上没有受过特殊训练的科学工作者或者将要作科学工作的学生而写的。希望提供数理统计的基本概念和方法,并指出它们的一般性的应用。大多数的例子选自生物学以及与之有关的学科,这是因为我们对数理统计在这些学科中的应用比较熟悉的缘故。介绍的次序是使方差分析法尽可能现实地提前介绍。

R. G. D. 斯蒂尔

J. H. 托里

符号集注

符号	意 义
$\neq$	不等于, 如 $3 \neq 4$
$>$	大于, 如 $5 > 2$
$\geq$	大于或等于
$<$	小于, 如 $3 < 7$
$\leq$	小于或等于
$ $	绝对值, 如 $ -7 = 7$
Σ	求和
$\cdots$	代表一组显然省写的量, 如 $1, \cdots, 10$
$n!$	$n(n-1)\cdots 1$ 并称为 n 的阶乘, 如 $3! = 3(2)1 = 6$
$-$	上杠, 用来表示一个算术平均数
$\wedge$	帽号, 用来表示一个估计量而不是真值; 多数加在希腊字母上
希腊字母	除少数例外, 指总体参量
μ	总体平均数
σ^2, σ	总体方差, 总体标准离差
τ, β , 等	总体平均数的分量; 通常用在线性模型中
ε	一个真正的实验误差
δ	一个真正的抽样误差
β	总体回归系数
ρ	总体相关系数
N, S^2	这些拉丁字母在第 22 章用作为总体的符号

以上的希腊字母在需予澄清的情况下也可以加上添标。例

如:

$\mu_{\bar{x}}$ $\bar{x}$ 的一个总体的平均数

$\beta_{yz \cdot x}$	为固定的 Z, Y 在 X 上的回归
τ_i	对接受第 i 种处理的总体平均数的作用分量
不用作为总体参量的希腊字母为:	
α	犯 I 型误差的概率
$1-\alpha$	置信系数
β	犯 II 型误差的概率
$1-\beta$	一项数理统计检验的辨别能力
χ^2	一个常见的检验判据
拉丁字母	用作为一般性的符号, 包括作为样本统计量的符号
X	一个变量(也常使用 Y)
X_i, X_{ij}	个别观测数据
$X_{i.}, X_{..}$	观测数据的总计数
x_i	$X_i - \bar{x}$
x'_i	$(X_i - \bar{x})/s$
D_j	成对观测数据间的差数
$n, n_{..}$	总的样本含量
n_{ij}	在第 i, j 格子内观测数据的数目
$\bar{x}, \bar{x}_{..}, \bar{x}_{i.}$	样本平均数, 整个或部分样本的平均数
$\bar{\bar{x}}$	一个平均数样本的平均数
$\bar{d}$	$\bar{x}_1 - \bar{x}_2$
s^2, s_x^2, s_y^2	样本方差, $\sigma^2, \sigma_x^2, \sigma_y^2$ 的无偏估计量
s, s_x, s_d	样本标准离差
$s_{y \cdot x}^2, s_{y \cdot 1 \dots k}^2$	为回归调整的样本方差
CL, CI	置信界限, 置信区间
l_1, l_2	置信界限的两端值
b	样本回归系数
$b_{y1, 2 \dots k}$	样本偏回归系数
b'	标准回归系数
r	样本总的或简单的相关系数

$r_{12 \cdot 3 \dots k}$	X_1 与 X_2 的样本偏相关系数
$R_{1 \cdot 2 \dots k}$	在 X_1 与其它变量间的多重相关系数
df, f	自由度
C, CT	校正项
SS	$\sum x_i^2$, 平方之和
MS	均方
E_a, E_b	为裂区设计用的误差项均方
E_{yy}, E_{xy}, E_{xx}	在协方差中误差项积数之和 (为其它变差来源使用相应的其它字母)
*	是显著的, 如 2.3*
**	是高度显著的, 如 14.37**
ns	是不显著的
lsd	最小显著差数
RE	相对效率
CV	变差系数 $(s/\bar{x})100$
$Q = \sum c_i T_i$	一个比较, 其中 c_i 是个常数, T_i 是个处理平均数, 且 $\sum c_i = 0$
fpc	有限总体校正数
psu	初级抽样单元
st	分为层次的, 用在添标上
K	$\sum c_i^2$
P	概率
$p, 1-p$	在二项分布中的概率
z, t, F	常见的检验判据
H_0	解消假设
H_1	备择假设, 通常代表一个备择假设集
∞	无穷大

目 录

序言	xi
符号集注	xv
第一章 绪论	1
1.1 数理统计的定义	1
1.2 数理统计的简史	2
1.3 数理统计与科学方法	4
1.4 学习数理统计	6
第二章 观测数据	9
2.1 引言	9
2.2 变量	9
2.3 分布	10
2.4 总体与样本	11
2.5 随机样本,数据的搜集	12
2.6 数据的表示、综合与表征	13
2.7 中心趋向的度量	16
2.8 散开度的度量	19
2.9 平均数的标准离差	23
2.10 变异性的系数或变差系数	25
2.11 一个例子	25
2.12 线性可加模型	26
2.13 置信推断	27
2.14 一个例子	29
2.15 在统计量的计算中使用编码法	31
2.16 频数表	32
2.17 一个例子	33

2.18	从频数表计算平均数和标准离差	34
2.19	频数表的作图表示法	36
2.20	有效数字	36
第三章	概率	38
3.1	引言	38
3.2	一些基本的概率	38
3.3	概率分布	40
3.4	正态分布	43
3.5	正态分布的概率; 概率表的使用	44
3.6	具平均数 μ 与方差 σ^2 的正态分布	47
3.7	平均数的分布	49
3.8	χ^2 分布	50
3.9	Student t 分布	52
3.10	估计与推断	54
3.11	样本结果的预测	58
第四章	从一个正态分布抽样	60
4.1	引言	60
4.2	一个正态分布的总体	60
4.3	来自正态分布的随机样本	62
4.4	样本平均数的分布	64
4.5	样本方差与标准离差的分布	67
4.6	s^2 的无偏倚性	68
4.7	平均数的标准离差或标准误差	68
4.8	Student t 的分布	69
4.9	置信推断	71
4.10	差数的抽样	73
4.11	抽样的总结	78
4.12	假设的检验	79
第五章	两个样本平均数间的比较	83
5.1	引言	83

5.2	显著性的检验	83
5.3	对两个或多于两个平均数的检验的基础	89
5.4	数据不成对但等方差的两个样本平均数的比较	90
5.5	线性可加模型	95
5.6	样本平均数的比较; 成对数据	96
5.7	成对比较的线性可加模型	99
5.8	不成对的数据与不等的方差	100
5.9	方差的齐性假设的检验	101
5.10	两平均数间差数的置信区间	104
5.11	样本含量与差数的辨别	104
5.12	Stein 两阶段样本	107
第六章	实验设计的原理	109
6.1	引言	109
6.2	什么是实验	109
6.3	实验的目的	110
6.4	实验单元与处理	111
6.5	实验误差	112
6.6	重复及其作用	112
6.7	影响于重复数目的因素	114
6.8	含有少数几种处理的设计的相对精确度	115
6.9	误差控制	116
6.10	处理的选择	119
6.11	技术上的改进	120
6.12	随机化	121
6.13	数理统计推断	122
第七章	方差分析 I: 单向分类	124
7.1	引言	124
7.2	完全随机设计	124
7.3	单种分类判据的数据、有等重复的任何组群数 的方差分析	126

7.4	最小显著差数	132
7.5	Duncan 新的多范围检验法	134
7.6	Tukey 的 w 法	137
7.7	Student-Newman-Keul 的检验法	138
7.8	全部平均数跟对照作比较	139
7.9	具单种分类判据的数据, 不等重复的任何数目组 群用的方差分析	140
7.10	线性可加模型	143
7.11	次级样本的方差分析, 等次级样本数	148
7.12	次级样本的线性模型	153
7.13	次级样本的方差分析, 不等的次级样本数	155
7.14	在计划含有不等次级样本的实验中的方差分量	158
7.15	方差分析内在的假定	159
第八章	方差分析 II: 多向分类	164
8.1	引言	164
8.2	随机化完全区组设计	164
8.3	任何处理数的方差分析; 随机化完全区组设计	166
8.4	误差项的性质	171
8.5	漏失数据	173
8.6	在效率上得益的估计	176
8.7	随机化完全区组设计: 每实验单元有多个观测数 据	177
8.8	线性模型以及方差分析	180
8.9	双组群: 拉丁方	181
8.10	拉丁方的方差分析	183
8.11	在拉丁方中的漏失数据	187
8.12	在效率上得益的估计	189
8.13	拉丁方的线性模型	190
8.14	一个实验的样本含量	191
8.15	变换	194

第九章 直线回归	200
9.1 引言	200
9.2 Y 在 X 上的直线回归	200
9.3 直线回归模型及其理解	204
9.4 在直线回归中的假定与性质	205
9.5 在直线回归中变差的来源	207
9.6 回归的与调整的值	208
9.7 标准离差、置信区间及假设的检验	210
9.8 用协变观测数据控制变差	213
9.9 在两个回归间的差异	214
9.10 一项预计及其方差	217
9.11 X 的预计, 模型 I	219
9.12 双变量分布, 模型 II	219
9.13 通过原点的回归	221
9.14 加权的回归分析	223
第十章 线性相关	226
10.1 引言	226
10.2 相关与相关系数	226
10.3 相关与回归	231
10.4 抽样分布, 置信区间, 假设的检验	232
10.5 相关系数的齐性	234
10.6 组内相关	236
第十一章 方差分析 III: 析因实验	238
11.1 引言	238
11.2 析因实验	238
11.3 2×2 析因实验, 一个例子	245
11.4 $3 \times 3 \times 2$ 或 $3^2 \times 2$ 析因实验, 一个例子	250
11.5 析因实验的线性模型	257
11.6 单个自由度比较	263
11.7 n 向分类与析因实验; 响应面	269

11.8	个别自由度; 等间距处理	272
11.9	为不可加性的一单个自由度	279
第十二章	方差分析 IV: 裂区实验与分析	284
12.1	引言	284
12.2	裂区设计	284
12.3	裂区设计的一例	289
12.4	在裂区设计中的漏失数据	294
12.5	时间裂区	296
12.6	裂区模型	300
12.7	时空裂区	301
12.8	相似实验的系列	303
第十三章	方差分析 V: 不等次级组类含量	308
13.1	引言	308
13.2	不成比例次级组类含量; 概论	308
13.3	不成比例次级组类含量; 安配常数的方法	314
13.4	不成比例次级组类含量; 平均数加权平方的方法	326
13.5	不成比例次级组类含量; 为 $r \times 2$ 表用的方法	329
13.6	不成比例次级组类含量; 2×2 表	333
第十四章	多重的与偏的回归和相关	339
14.1	引言	339
14.2	线性方程以及它在多于两维时的理解	340
14.3	偏、总、与多重线性回归	341
14.4	样本多重线性回归方程	343
14.5	多重线性回归方程; 三变量	344
14.6	标准偏回归系数	347
14.7	偏相关系数与多重相关系数; 三个变量	348
14.8	显著性的检验; 三个变量	350
14.9	多重线性回归; 多于三个变量的算法	353
14.10	简易 Doolittle 法	354

14.11	多重回归的显著性检验	362
14.12	偏回归系数的标准误差及其显著性的检验	363
14.13	标准偏回归系数	365
14.14	剔除或外加一个自变量	365
14.15	偏相关	368
第十五章	协方差分析	372
15.1	引言	372
15.2	协方差分析的用处	372
15.3	协方差的模型与假定	376
15.4	检验调整的处理平均数	378
15.5	在随机化完全区组设计中的协方差	380
15.6	处理平均数的调整	385
15.7	协方差导致的精确度的增加	387
15.8	协方差的划分	388
15.9	回归系数的齐性	390
15.10	把处理项平方之和划分了的协方差	392
15.11	用协方差作漏失数据的估计	396
15.12	具两个协变量的协方差	398
第十六章	曲线回归	407
16.1	引言	407
16.2	曲线回归	407
16.3	对数或指数曲线	408
16.4	二次多项式	414
16.5	二次多项式, 一个例子	416
16.6	高次多项式	418
16.7	多项式与协方差	418
16.8	正交多项式	418
第十七章	χ^2 的一些应用	424
17.1	引言	424
17.2	σ^2 的置信区间	424