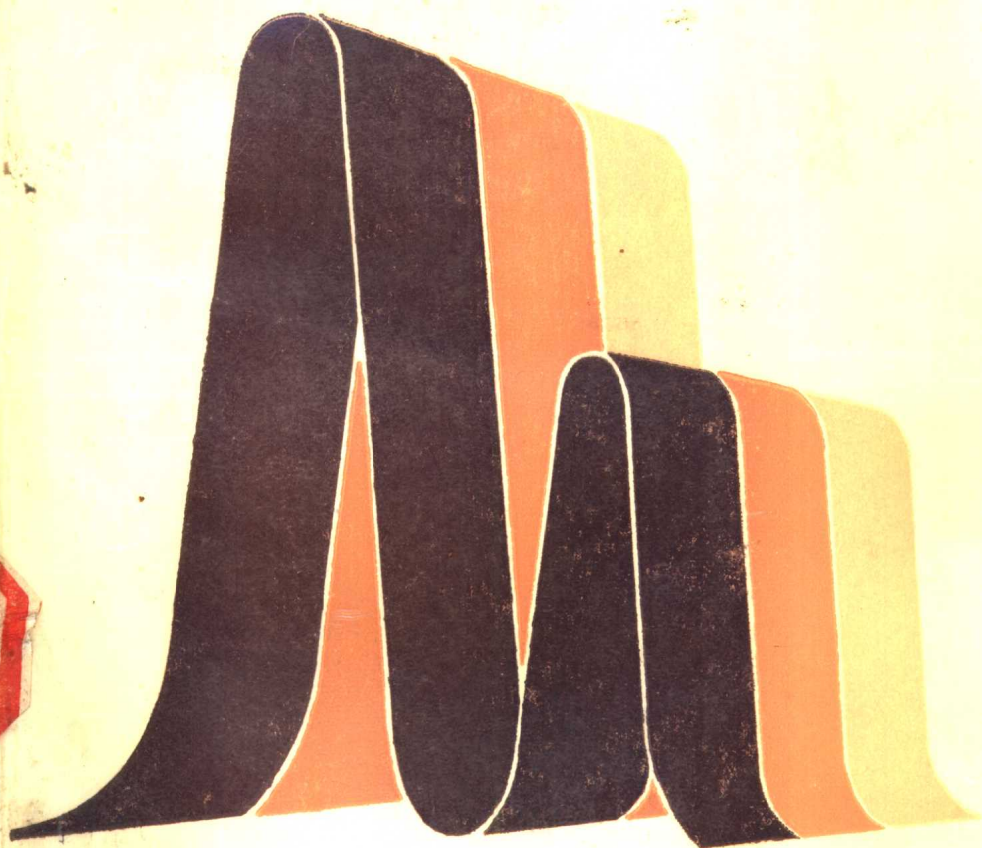


Modern Probability & Statistics

现代概率统计丛书

〔美〕G·W·斯内德克 等著

统计方法



上海翻译出版公司

统 计 方 法

GEORGE W. SNEDECOR
[美] WILLIAM G. COCHRAN

刘垂玕 杨义群 张抱膝 译

上海翻译出版公司

1990

G. W. Snedecor

W. G. Cochran

Statistical Methods

Iowa State Univ. Press

1982

统计方法

G. W. Snedecor 著
W. G. Cochran

刘垂珩 杨义群 张抱膝 译

上海翻译出版公司

(上海复兴中路597号)

邮政编码: 200020

本书在上海发行所发行 常熟高专印刷厂印刷

开本 850×1168 1/32 印张 22.5 字数 606,900

1991年4月第1版 1991年4月第1次印刷

印数 1—1,500

ISBN7-80514-291-2/O·77 定价: 17.00元

译者的话

由George W. 斯内德克与William G. 科克伦合著的《统计方法》一书，是统计学的经典名著之一。这本书的初版是由斯内德克单独在1937年完成，并在1938，1940，1946，1956年一再重版而出到第五版。在此以后，科克伦远渡重洋，离开了英国的Rothamsted实验站来到美国的Iowa州立大学与斯内德克共事。他们两人濡沫切磋，相得益彰。1967年本书的第六版即是由他们两人共同完成的。七年以后，斯内德克仙去，但科克伦逾花甲之年而雄心不减，仍致力于使本书更臻完美。就在这项工作即将完成之际，科克伦本人又在1980年3月29日以七十一岁的高龄溘然长逝。所幸的是，科克伦未竟之业，立即由Iowa州立大学的David F. 考克斯接了下来。考克斯整理了科克伦的遗稿，在当年（1980）就出了本书的第七版，也就是译者据以译成中文的这个版本。这个版本极受欢迎，不得不在1982年再次印刷。

一本科学著作，能够经历五十个春秋，再版而七，实不多见。

统计学虽应用极为广泛，但其理论上的深奥却常使初学者望而生畏。本书在处理实用性与科学性方面的深思熟虑，不愧为深入浅出的典范。科学地阐明统计方法而又不拘泥于数学形式的严谨，使具有不太多数学基础知识的读者能不背离科学原则地掌握及运用统计方法，这是本书独到之处，可能也就是其强大生命力之所在。以斯内德克及科克伦的数学功底，原来可以把本书写得象他们的其它学术著述那样抽象严谨简练，然而他们推出了现在这样的一本书。其实，也只有登凌绝顶高，才能一览众山小。译者不能不叹服两位大师的匠心，留下了这本经半个世纪的锤炼

而成的名著。

然而，科克伦继斯内德克而逝的影响，在本书中仍多处可见。考克斯虽整理了科克伦的手稿，但他并不能代替科克伦来思考。本书中有不少的错误，可能就是科克伦未能进一步校审而遗留下来的痕迹。译者尽自己的水平，在发现这些痕迹时，通过译者注来加以指出。

由于斯内德克及科克伦都已不在人世，本书所据的当是最后一个版本。译者虽尽力而为，但未必能全如人意。译文中有不当之处，还请读者给予赐正。

最后，译者感谢上海翻译出版公司的大力支持，使得本书得以尽快地与我国读者见面。

安徽农学院	刘垂玕
浙江农业大学	杨义群
南京大学	张抱膝

一九八七年春

原 著 者 简 介

W. G. 科克伦 (William G. Cochran) 在他1980年逝世时, 是哈佛大学的统计学荣誉教授。以前, 他曾在约翰·霍普金斯大学、北卡罗来纳州立大学以及衣阿华州立大学任教。他是国家科学院及美国人文及自然科学学院的院士。格拉斯哥及约翰·霍普金斯大学曾授予他名誉法学博士的称号。他曾任国际统计学研究所、美国统计学会、数学统计学研究所及生物统计学会的主任或主席。他的著作包括在专业刊物上发表的许多研究论文, 以及专著: 《抽样技术》(1977年第三版, 有中译本, 由张尧庭等译) 及《试验设计》(与G. M. Cox合著, 1957年第二版)。

G. W. 斯内德克 (George W. Snedecor) 在他1974年逝世以前一直是衣阿华州立大学的统计学荣誉教授, 在那里从1913年到1958年他一直在讲坛上, 并且担任了统计学实验室主任达十四年之久。他的著作甚多, 除去论文及研究报告外, 还有专著《相关与机器计算》(与H. A. Wallace合著) 及《方差、协方差分析的计算与解释》。他从密执安大学获得理学硕士学位, 并从北卡罗来纳州立大学及衣阿华州立大学获得名誉理学博士学位。他是国际统计学研究所理事, 曾任美国统计学会主席, 还是英国皇家统计学会的名誉会员。他还是加州圣迭戈的美国海军电子实验室的顾问。

原著者序

本书的前几版曾被广泛用作统计学导论的教科书以及需要统计技术去收集与分析数据的工作者的参考书。作为教科书，本书所含丰富材料超过了通常学时上的安排，在序言之前，列出了作为一学期课程内容的建议。本书大部分内容所需的数学水平几乎不超过初等代数。

本版有所重写并重新加以安排，为的是使本书更易于应用并使某些章更短些。现在的第一章介绍了本书主要论述的资料类型——抽样调查、受控试验以及比较观测研究——加上随机抽样的概念。接下去的几章为频数分布、平均数及标准差、正态分布以及假设检验。

再后面的几章，除去把方差分析的随机效应模型以及假定的欠周的内容分别单列成章外，其它就跟第六版中的通用教程一样。

为拟合多重线性回归的计算已改为利用除换运算 (sweep operator)。这是一种常规方法，易于学习，并易于在电子计算机上编成程序。还有，若用虚变量 X 去表示分类，方差与协方差分析的计算就可以通过拟合多重线性回归的电算程序来完成。教学上不可能把标准程序语言的学用列入统计学的简明教程。目前最合适的方法或许就是为那些未曾学过程序语言的学生建立可以计算这本简明教程中的数值实例的简单程序。本版中未列入任何程序，但不时指出计算机怎样表示出结果以及在计算机输出的要求下进行怎样的运算。

在本版中首次出现的题目有：

概率纸 (节 2.6)*。

*此处应为节 4.13。——译者。

一系列实验中至少有一次成功的概率(节7.5)。

估计方差集合相等的 Levene 稳健检验(节13.11)。

重复测量试验给出的处理中秩序的平衡(节14.9)。

方差分析中变换的不同效应的联合研究(节15.15)。

析因试验的 Yates 算法(节16.12)。

重复测量的试验(节16.16)。

多重回归中挑选预测变量子集的 Mallows C_p 判据(节17.14)。

抽样调查中的非抽样误差(节21.13)。

增加了一些探测离群值(outlier)的表格:

表A16在5%及1%水平上给出了正态样本的最大赋范残差
 $\max |X - \bar{X}| / \sqrt{\sum (X - \bar{X})^2}$ 以及独立正态的 d 的有关量
 $\max |d| / \sqrt{\sum d^2}$ 。

表A18给出了两向分类中最大赋范残差的显著水平。

表A17给出了离群值快速检验的 Dixon 判据。

另外,以前的一些表格已予扩充。

序言后面,列出了本书常用的符号。

我们感谢允许复制他们图表资料的作者及出版者^{*}。感谢使用过本书前几版而给予指教的教师们。

William G. 科伦伦

^{*}具体人名及出版社名略。——译者。

常用符号

一般:

希腊字母 $\alpha, \beta, \dots$: 总体参数

拉丁字母 $a, b, \dots$ 或符号 $\hat{\alpha}, \hat{\beta}, \dots$: 样本参数

X, Y : 观测值

μ_X, μ_Y : 总体平均数; $\bar{X}, \bar{Y}$: 样本平均数

$x = X - \bar{X}$: 对样本平均数的离差

n : 样本容量

a 峰态检验中 (均离差) / (标准差) = $\sum |x| / (ns)$

a_i, A_i 第 i 组的固定效应及随机效应

$[A]$ 因子 A 的总析因效应

β_i, b_i 多重回归中 Y 对 X_i 的回归系数

$\sqrt{\beta_1} = \gamma_1$ $E(X - \mu)^3 / \sigma^2$, 偏度的测量

$\sqrt{b_1} = g_1$ $m_3 / m_2^{3/2}$, 测量偏度的样本估计值

c_{ij} 用于多重回归的乘子: $c_{ij}\sigma_e^2 = \text{Cov}(b_i, b_j)$

C_p 挑选预测变量集合的 Mallows 判据

χ^2 卡方

cv 变异系数: σ/μ 或 $s/\bar{X}$

D_i $X_{1i} - X_{2i}$, 对于成对样本, $i = 1, 2, \dots, n$

$d_{y..x}$ Y 对于 X 的样本回归的离差

$d_{y..12\dots p}$ Y 对于 $X_1, X_2, \dots, X_p$ 的样本回归的离差

E 期望值: $E(X) = X$ 的期望值 (均值)

e 模型的随机残差

e X 的测量误差; 也表示自然对数的底, 即

$$e = 2.71828 \dots$$

f	某组的观测频数
$f(X)$	X 的函数; 例如 X^2 、 e^X 、 $\log X$
F	期望频数
F	方差比, s_1^2/s_2^2
γ_2, g_2	峰态的测量: $\gamma_2 = E(X - \mu)^4/\sigma^4 - 3$; $g_2 = m^4/m_2^2 - 3$
I	组区间距
L	线性函数: $\sum \lambda_i X_i$ 或 $\sum L_i X_i$, 其中 λ_i , L_i 是数字
$m_1, \dots, m_4$	前四阶样本矩: $m_i = \Sigma(X - \bar{X})^i/n$
n_h	调和平均数
N	有限总体的容量
ν	自由度的数目
P	概率
p	二项总体中成功的概率: $q = 1 - p$
Q	学生氏化的范围, (样本范围)/ s
ϕ	n/N , 抽样分数
ρ, r	Y 与 X 间的相关系数
ρ_I, r_I	组内相关系数
r_s	Spearman 的秩相关系数
R^2	多重相关系数的平方
R_p	由 $X_1, X_2, \dots, X_p$ 上回归而对 Σy^2 的缩减
σ, s	标准差
$\sigma_{y \cdot x}, s_{y \cdot x}$	Y 对 X 回归的残余的标准差
$\sigma_{\bar{X}}, s_{\bar{X}}$	样本平均数的标准误差
Σ	和, $\sum_{i=1}^n X_i = X_1 + X_2 + \dots + X_n$
	学生氏的 $t = (X - \mu)/s$

U	X 的码值: $(X - X_0)/I$
V	方差
w	抽样范围, 加权平方和中的权
W_h	分层抽样中 h 层的权: $W_h = N_h/N$
X_0	加码(U)标度为零处的 X 的值
Z	$(X - \mu)/\sigma$ 观察 X 的标准表达形式

缩语:

AH	备择假设
ANOVA	方差分析
$\text{Cov}(XY)$	X 与 Y 的协方差 $= E(X - \mu_X)(Y - \mu_Y)$
CR	完全随机化
deff	设计效应
df	自由度
ESD	极端学生氏化离差
H_0, H_A	原假设, 备择假设
$\ln$	自然对数
LSD	最小显著差异
MLR	多重线性回归
MNR	最大赋范残差
MS	均方
MSE	均方误差
NH	原假设
RB	随机化区组
SD	标准差
SE	标准误差
SS	平方和

目 录

译者的话	(1)
原著者简介	(3)
原著者序	(4)
常用符号	(6)
第一章 绪论	(1)
1.1 引言	(1)
1.2 本章的目的	(2)
1.3 抽样调查的例子	(2)
1.4 抽样的问题	(3)
1.5 有偏抽样	(4)
1.6 随机抽样	(5)
1.7 随机数字表	(7)
1.8 估计值的抽样分布	(8)
1.9 不同剂型比较效应的研究	(11)
1.10 从比较研究得出结论的问题	(12)
1.11 小儿麻痹症疫苗的有效性	(13)
1.12 不同吸烟组的死亡率	(15)
1.13 观测研究	(15)
1.14 概要	(16)
术语	(18)
参考文献	(19)
第二章 频数分布	(21)
2.1 定量数据	(21)
2.2 频数分布	(21)

2.3	分组的频数分布	(22)
2.4	组限	(25)
2.5	累积频数分布	(26)
2.6	概率分布	(29)
	术语	(32)
	参考文献	(32)
第三章	平均数与标准差	(33)
3.1	算术平均数	(33)
3.2	总体平均数	(35)
3.3	总体标准差	(37)
3.4	两状态总体	(38)
3.5	样本标准差	(39)
3.6	关于计算机的注记	(43)
3.7	用频数分布计算 $\bar{X}$ 和 s	(44)
3.8	数值例子	(45)
3.9	变异系数	(48)
	术语	(49)
	参考文献	(50)
第四章	正态分布	(51)
4.1	正态分布的总体	(51)
4.2	采用正态分布的原因	(53)
4.3	正态分布表	(53)
4.4	样本平均数的标准差	(57)
4.5	样本平均数的频率分布	(59)
4.6	三个实例	(59)
4.7	σ 已知时 μ 的置信限	(65)
4.8	样本容量	(68)
4.9	学生氏 t 分布	(70)
4.10	μ 基于 t 分布的置信限	(72)

4.11	t 分布的试验抽样	(74)
4.12	置信区间的样本验证	(76)
4.13	概率纸	(76)
4.14	模拟正态的有限总体	(80)
	术语	(81)
	参考文献	(82)
第五章	假设检验	(83)
5.1	引言	(83)
5.2	正态总体的均值检验 (σ 已知)	(83)
5.3	显著性检验和置信区间	(85)
5.4	显著性检验的实际应用	(86)
5.5	单边或单尾检验	(87)
5.6	显著性检验的能力	(89)
5.7	σ 未知时, 平均数的检验	(92)
5.8	其他显著性检验	(95)
5.9	s^2 的频数分布	(96)
5.10	σ^2 的区间估计	(97)
5.11	关于 σ^2 值的原假设的检验	(98)
5.12	χ^2 拟合优度检验	(99)
5.13	偏斜度检验	(103)
5.14	峰度检验	(105)
	术语	(107)
	参考文献	(107)
第六章	两个样本的比较	(109)
6.1	差值的估计量与检验	(109)
6.2	模拟的配对试验	(110)
6.3	配对试验的例子	(113)
6.4	配对的条件	(115)
6.5	两独立样本平均数的比较	(116)

6.6	差值的方差·····	(117)
6.7	方差的合并估计量·····	(118)
6.8	比较两个相同容量的样本的试验·····	(119)
6.9	不等容量的样本·····	(121)
6.10	防偏措施: 随机化·····	(124)
6.11	$\sigma_1 \neq \sigma_2$ 时独立样本的分析·····	(126)
6.12	两方差齐性的检验·····	(128)
6.13	配对样本与独立样本·····	(130)
6.14	比较试验中的样本容量·····	(134)
	术语·····	(138)
	参考文献·····	(138)
第七章	二项分布 ·····	(140)
7.1	引言·····	(140)
7.2	一些简单的概率法则·····	(140)
7.3	二项分布·····	(145)
7.4	对二项分布抽样·····	(147)
7.5	至少有一次成功的概率·····	(150)
7.6	正态近似及连续性校正·····	(153)
7.7	二项比例的显著性检验·····	(157)
7.8	二项比例的置信限·····	(158)
7.9	配对样本的比例比较·····	(159)
7.10	独立样本的比例比较: 2×2 表·····	(162)
7.11	2×2 联列表中的 χ^2 检验·····	(165)
7.12	两属性的独立性检验·····	(168)
7.13	比较两个比例所需的样本容量·····	(169)
7.14	Poisson 分布·····	(171)
	术语·····	(175)
	参考文献·····	(175)

第八章 简捷法与非参数法	(177)
8.1 引言	(177)
8.2 样本中位数	(177)
8.3 由样本极差估计 σ	(180)
8.4 符号检验	(181)
8.5 基于极差的 t 检验替代	(184)
8.6 符号秩检验	(185)
8.7 两独立样本的 Lord 检验	(187)
8.8 两独立样本的秩和检验	(189)
8.9 秩检验与正态检验的比较	(191)
8.10 取有限个值的离散型标量: 随机化检验	(192)
术语	(195)
参考文献	(195)
第九章 回归	(197)
9.1 引言	(197)
9.2 拟合线性回归的计算	(197)
9.3 线性回归的数学模型	(202)
9.4 线性回归的方差分析	(207)
9.5 最小二乘法	(209)
9.6 观测研究中的回归	(211)
9.7 数值例子	(213)
9.8 由 X 预报 Y	(216)
9.9 预报总体回归线	(217)
9.10 预报新的个体 Y	(219)
9.11 对可疑大离差的检验	(221)
9.12 由 Y 预报 X : 线性校准	(222)
9.13 Galton 使用的“回归”术语	(225)
9.14 X 有误差时的回归	(225)
9.15 拟合通过原点的直线	(227)

术语	(229)
参考文献	(230)
第十章 相 关	(231)
10.1 样本相关系数 r	(231)
10.2 r 的性质	(233)
10.3 二元正态分布	(235)
10.4 相关系数的一些应用	(238)
10.5 原假设: $\rho=0$ 的 检 验	(244)
10.6 ρ 的置信限与假设检验	(245)
10.7 线性函数的方差	(248)
10.8 配对样本中两个相关方差的比较	(251)
10.9 非参数法: 秩相关	(253)
术语	(255)
参考文献	(255)
第十一章 单向分类与双向分类的频数分析	(257)
11.1 引言	(257)
11.2 有两类以上的单向分类	(257)
11.3 具有相同期望值的单向分类	(261)
11.4 检验 Poisson 总体具有相同平均数	(262)
11.5 附加检验	(263)
11.6 双向分类: $2 \times C$ 列联表	(265)
11.7 二项总体的齐性检验	(267)
11.8 有序分类	(270)
11.9 检验各项比例中的线性趋势	(274)
11.10 $R \times C$ 列联表	(276)
11.11 若干 2×2 表	(280)
术语	(284)
参考文献	(284)
第十二章 单向分类, 方差分析	(286)