

应用统计手册

YING YONG TONG JI SHOU CE

罗札。塞克斯著、罗永泰、史道济译

1	2	3	4	5	6	7
---	---	---	---	---	---	---

10	9	8	7	6	5	4
----	---	---	---	---	---	---

1	2	3	4	5	6	7
---	---	---	---	---	---	---

10	9	8	7	6	5	4
----	---	---	---	---	---	---

1	2	3	4	5	6	7
---	---	---	---	---	---	---

10	9	8	7	6	5	4
----	---	---	---	---	---	---

1	2	3	4	5	6	7
---	---	---	---	---	---	---

10	天津	科技	翻译	出版	公司	4
----	----	----	----	----	----	---

51.73673
658

3165155
应用统计手册

罗札 · 塞克斯 著

罗永泰 史道济 译

马逢时 李宁生 校



内 容 简 介

本书介绍了各种统计方法：主要包括单个总体、两个总体、多个总体的各种参数、非参数的估计、检验，相关与回归，方差分析，列联表分析等内容；给出了440个各种学科的统计例子；提供了利用统计方法解决实际问题的途径。书中还给出了各种统计用表及丰富的参考文献。本书是广大工程技术人员、自然科学工作者、医学、管理、经济等部门的有关人员，研究人员及大专院校师生必备的工具书。

序

Lothar Sachs
Applied Statistics
A Handbook of Techniques

应 用 统 计 手 册

罗扎·塞克斯著
罗永泰 史道济译
责任编辑 李丕章

天津科技翻译出版公司出版
天津市河西区吴家窑大街22号
新华书店天津发行所发行
河北省三河振兴印刷厂排版印刷

1988年5月第一版 1988年5月第一次印刷
开本：850×1168·1/32 印张：27.7
字数：693.8千 印数：0001—9000册
定价：8.25元
ISBN T-5433-0004-4/TB·1

译者说明

当今，数理统计在国民经济和科学研究中的应用越来越广泛深入。正如本书作者在第一版序言中所说的：我们的时代正是以概率论和数理统计的迅速发展，及其在科学、技术、经济和政策方面的应用为标志的。西德 Kiel 大学 Lothar Sachs 所著的《应用统计手册》是国际上一本著名手册，现已译成英文、俄文、西班牙文等文本出版。由于目前我国尚缺少一本较为系统、全面而又深入浅出地讲述统计方法的工具书，因此我们翻译出版了这本书，奉献给全国广大读者。

本书内容广泛、丰富，又不要求读者具有较高深数学知识。对于广大的医务、管理、科技工作者和高等院校的师生，本书既是一本可供学习的入门读物，又是一本十分有价值的参考手册。本书介绍了各种最新的统计方法和直至1984年的最新成果，提供了一般统计书籍中所没有的各种统计用表，以及大量的参考文献，因此它将是概率统计工作者所必备的工具书。

本书是根据 Springer-Verlag 公司出版的 New York 大学 Zenon Reynarowych 的英文版第二版 (1984年) “Applied Statistics, A Handbook of Techniques” 译出的。全书的翻译工作在马逢时教授的主持下进行，罗永泰、史道济合译第1章，罗永泰译第2、3、4章，史道济译第5、6、7章，马逢时教授、李宁生副教授负责全书的统一、审校和定稿，陈正辉副教授参加了第5、6章的审校工作。由于篇幅限制，我们删去了第0章初等数学预备知识(包括两张对数表)，也删去了最后的作者索引和主题索引，并将原书扉页上的表移到最后。

由于译者水平有限，译文中的缺点、错误以及不妥之处在所难免，敬请广大读者批评指正。

译者 1987年6月于天津

希 腊 字 母 表

希腊字母		英文读音	希腊字母	英文读音
A	α	Alpha	N	ν Nu
B	β	Beta	Ξ	ξ Xi
Γ	γ	Gamma	O	o Omicron
Δ	δ	Delta	Π	π Pi
E	ϵ	Epsilon	P	ρ Rho
Z	ζ	Zeta	Σ	σ Sigma
H	η	Eta	T	τ Tau
H	θ	Theta	Y	υ Upsilon
I	ι	Iota	Φ	φ Phi
K	κ	Kappa	X	χ Chi
Λ	λ	Lambda	Ψ	ψ Psi
M	μ	Mu	Ω	ω Omega

英文第二版序言

为了适应应用统计学的日益迅速发展，我们这一次新版的目
的如同第一版一样，对适用于广阔领域的不同应用统计分支，给出
了更深刻、更明确的简要说明和概括。本版在内容上进行了增加和
修订，加多了主题索引，参考文献也更加近代化和更加丰富了。
为了有利于对这一领域还陌生的读者，我已试图通过引用一些容
易理解，数学性不强，对于非统计学家更易读的书和较早期的一
些论文，而没有一味引用较新的论文。对于后面的情况，文中以
一种简便的方式来原因所引用的论文。例如，用“（也参
阅…）”。

我向许多读者表示由衷的感谢，他们所提出的问题和建
议有助于这次新版。我非常严肃认真地采纳了许多建议，但也并不是
完全照搬他们的观点。对于任何有益的建议，我都热忱地欢迎。
我对于Springer—Verlag New York 的职员们表示特别的感
谢！

Klausdorf/Schwentine

LOTHAR SACHS

英文第一版序言

现在英文译本与俄文、西班牙文版一起出版了，这是根据本
书最新的德文第5版修订版译出的。由于本书提供的纵横交叉的
参考文献非常容易学习和应用，所以它已成为在工程、医学和管
理科学等领域的科学家非常欢迎的学习和参考资料。由于本书在
正文和目录中提供了许多使用方便的表、辅助表及其使用方法，
因而提高了它作为工具手册的价值。英文版比德文原版包括更多

的内容。我非常感谢大家在交谈中、通讯中所提出的启发性建议和对前几版的批评，並继续欢迎大家提出批评和建议。我特别感谢St. Paul, Minnesota的Dorothy Aeppli女士，在英文手稿的准备过程中，她提出了许多有价值的建议。作者和译者将对任何错误和缺点负责，並由衷欢迎所提出的任何改进意见。

我本人非常感谢译者Zenon Reynarowych 先生，他的精确而巧妙的翻译使本书变得更加清晰易懂，也非常感谢Springer—Verlag出版社。

LOTHAR SACHS

第一版序言（1967.11）

“这不仅仅是出于偶然”，大约250年以前伦敦的一位医生Arbuthnott这样想，当时他观察了八十年中按年度人口登记造册，发现男性的出生数总是高于女性的出生数。基于这样大小的样本，他的推断是相当可靠的。他可以在每一男性出生数后画一个“+”号（当然，这个数目要比女性出生数目高），这样可以建立一个符号检验。对于较大的样本，某种特殊符号的 $\frac{2}{3}$ 多数就可以充分满足检验的需要。当样本数较小时，一个 $\frac{4}{5}$ 或 $\frac{9}{10}$ 的多数也足以令人可靠地检验差异。

我们的时代正是以概率论和数理统计的迅速发展，以及它们在科学、技术、经济和政策方面的应用为标志的。

（致谢略）

第二版及第三版序言

在这一版里我们对某些章节进行了补充和修订，另一些章节

则完全重新撰写，特别是对于基于算术开方根的基本运算的章节以及有关统计学的基本任务，标准差及方差计算，第Ⅰ类风险和第Ⅱ类风险计算，对于 μ 已知或未知时 $\sigma=\sigma_0$ 的检验，利用反正弦变换对于 $\pi_1=\pi_2$ 及 $\pi_1-\pi_2=d_0$ 的检验，四格表的t检验，当第Ⅰ类及第Ⅱ类风险给定时这个检验所要求的样本大小，U检验，H检验以及中位数的置信区间，还有Sperman秩相关，点两列和多重相关，对于两个独立变量的线性回归多元方法以及试验设计和方差分析模型。对于下面一些图表进行了增补或完全修订：标准正态分布的临界值，t分布及 χ^2 分布的临界值，Hartley的 $F_{\max}$ 临界值，对于成对差的Wilcoxon的R值， $e^{-\lambda}$ 和 $\arcsin \sqrt{P}$ 的值，以及相关系数的z变换表和在单、双侧问题中 $\rho=0$ 检验的临界值表。也对书目进行了仔细检查。在第三版里，除了修正式、大量的简化式以及改进公式外，还加进了最新的材料。扩充了一些统计表（69a，80，84，98，99以及在4.5.1节和5.3.3节中未标号的表）。对于参考书目也进行完全的修订，作者索引也有新的特色，而且考虑了对于第一版和第二版的几乎所有建议。

第四版序言（1973.6）

这一修订版为本书起了一个更合适的书名。由于本书含有一系列的公式表，大量纵横交叉的参考文献，丰富的书目，以及作者索引和详细的主题索引，因而本书既可以作为一本入门读物和继续学习研究的教材，又可以作为一本有用的参考书。本版已使本书更加精炼简化，叙述更加准确。书目的大部分内容都根据最新所得到的材料进行了变更、代换，以至重新修订扩充并加入新内容。对许多统计表也作了同样的更动（扉页正面的索引以及表13，14，28，43，48，56，65，75，84，183，185和在

1.2.3, 1.6.4, 3.9.1节中未标号的表, 还有在书后面的表)。更主要的更动出现在作者所修订的书“统计方法”(“Statistical Methods, A Primer for Practitioners in Science, Medicine, Engineering, Economics, Psychology and Sociology”)中, 这本书可以作为一本便于迅速查阅的工具书。许多人所提供的建设性批评对于这两本书的写作都有很大帮助, 特别是工程师们提供的意见更是如此。本书对于医学院的学生也将是有益的。因为我已经包含了许多医学院在生物数学, 医学统计学所必需的材料。

第五版序言 (1978.7)

这一版使我能有机会引进一些简化和增补的内容, 更精确地用公式表示问题和解答。我非常感谢Clyde Y. Kramer教授, 使我能从他的书(A First Course in Methods of Multivariate Analysis, Virginia Polytechnic Institute and State University, Blacksburg, Virginia, 1972)中引用 Bonferroni χ^2 统计量的上界(附录D, 326—351)这些是由G. B. Beus和D. R. Jensen在1967. 9计算得到的。我非常欢迎任何批评和建议以便改进。

LOTHAR SACHS

符 号 说 明

$>$	大于
$\geq$	大于或等于
$\approx$	近似等于
$\neq$	不等于
Σ	和号； Σx 表示将所有 x 加起来
e	自然对数的底；一个常数，约为2.71828
P	概率
E	事件
X	随机变量，一个可能取指定集合中任意一个值的量；任何特殊的或个别的值，用 x 表示（例如 $X = \text{高度}$ ， $x = 173\text{cm}$ ）；如果对每个实数 x ，概率 $P(X \leq x)$ 存在，那末 X 称为随机变量〔因此 X, Y, Z 表示随机变量， x, y, z 表示它们所取的个别值〕；在本书中，我们几乎总是只用 x
∞	无穷
π	总体的相对频率
μ	总体的算术平均值
σ	总体的标准差
$\hat{p}$	样本的相对频率（用 $\hat{p}$ 估计 π ；估计值常用 $\hat{}$ 表示）
$\bar{x}$	样本（变量 X ）的算术平均
s	样本的标准差：标准差的平方 s^2 称为样本方差； σ^2 是总体的方差
n	样本大小，样本量，样本容量
N	总体的大小
k	组数

- z 标准正态变量, z 检验的统计量; z 检验是标准正态分布应用于有大样本假设的检验。对标准正态分布, 即均值为0, 方差为1的正态分布 (记作 $N(0,1)$), 我们用下列记号:
1. 纵坐标: $f(z)$, 例 $f(2.0)=0.054$ 或 0.0539910
 2. 累积分布函数: $F(z)$, 例 $F(2.0)=P(Z \leq 2.0)=0.977$ 或 0.9772499 ; $F(2.0)$ 是累积概率, 或正态概率函数从 $-\infty$ 到 $z=2.0$ 的积分
- f 记在单元中的频数
- V 变异系数
- $\sim_x$ 样本 (变量 X) 的中位数
- $s_{\bar{x}}$ 样本算术平均值的标准差
- $s_{\tilde{x}}$ 样本中位数的标准差
- R 极差=极端样本值之间的距离
- S 置信系数 ($S=1-\alpha$)
- α 显著性水平, 第 I 类风险, 拒绝有效零假设的小概率
- β 第 II 类风险, 保留无效零假设的概率
- H_0 零假设
- H_A 备择假设或对立假设
- z z 检验的临界值: z 是标准正态分布的上侧 α 百分位点 (横坐标值)。对这种检验或利用其它临界值的检验, 假如零假设为真, P 值给出了抽样结果的概率 (因此, 如果这个值非常小, 它并不是表示实际差别的大小)
- $\hat{t}$ t 检验的统计量; t 检验, 例如利用 t 分布或学生分布 (来自正态分布不大样本的分布律) 检验两个均值的相等性。
- v 或 DF , (分布的) 自由度
- $t_{v,\alpha}$ t 检验的临界值, 下标表示 t_v 分布的自由度 (v) 和百分位

点 (α)

χ^2 χ^2 检验的统计量; χ^2 检验, 例如检查观测的与理论的频数分布之间的差异, $\chi^2_{v, \alpha}$ 是 χ^2 检验的临界值

$\widehat{F}$ 方差比, F 检验的统计量; F 检验是利用 F 分布 (一种方差之商的理论分布) 检查两个方差之间的差异; $F_{\nu_1, \nu_2, \alpha}$ 是 F 检验的临界值

C_n^x 或 $\binom{n}{x}$, 二项系数, n 个元素中同时取出 x 个元素的组合数

$!$ 阶乘符号; n 个元素按一定顺序排列的种数为 $n! = n(n-1)(n-2) \times \cdots \times 3 \times 2 \times 1$

λ 参数, 为 Poisson 分布的均值与方差, Poisson 分布是在研究失效数据中有用的离散分布

CI 置信区间, 未知参数的范围, 具有这样一种性质的随机区间: 这个随机区间将以 $1-\alpha$ (例如 $1-0.05=0.95$) 的概率包含真实的未知参数, 例如 95% CI

MD (关于均值的) 平均偏差 = $(1/n) \sum |x - \bar{x}|$

Q 关于均值的平方和 (例如 $Q_x = \sum_{i=1}^{i=n} (x_i - \bar{x})^2 =$

$\sum (x - \bar{x})^2$)

U Wilcoxon-Mann-Whitney 检验的统计量: 两个独立样本的比较

H Kruskal-Wallis 检验的统计量: n 个独立样本的比较

O 观测的频数, 占有数

E 期望频数, 期望数

a, b, c, d 四格表的频数 (单元记录)

ρ 总体的相关系数: $-1 \leq \rho \leq 1$

r 样本的相关系数: $-1 \leq r \leq 1$

- β 总体的回归系数 (例如 β_{yx})
- b 样本的回归系数或斜率; 给出回归直线的方向; 常用二个下标, 例如 b_{yx} , 第二个下标表示自变量, 第一个下标表示由这个自变量的预测变量
- r_s 样本的Spearman秩相关系数: $-1 \leq r_s \leq 1$
- $s_{y \cdot x}$ 由X (样本) 估计Y的标准差
- s_a 截距的标准差
- s_b 斜率的标准差
- z 相关系数的正态化变换 (注意 z 上面的园点)
- E_{yx} y 对 x 的 (样本的) 相关比: 对检验回归的线性是重要的
- $r_{12 \cdot 3}$ 偏相关系数
- $R_{1 \cdot 23}$ 复相关系数
- SS 平方和, 例如 $\sum (x - \bar{x})^2$
- MS 均方: 样本方差 $s^2 = \sum (x - \bar{x})^2 / (n - 1)$ 是均方, 因为平方和被它的 $(n - 1)$ 个自由度相除
- MSE 误差的均方, 均方误差: 度量数据集的未解释的变异性, 並用作试验固有的随机变差的估计量; 它是试验误差的方差的无偏估计
- LSD 两个均值之间的最小显著差异
- SSA 因子A的平方和, 总变差中由因子A的水平的均值不同引起的那部分; 因子是一系列有关的处理或有关的分组
- MSA 因子A的均方: $MSA = SSA / (a - 1)$, 因子A的主效应均方
- SSAB 交互作用AB的平方和, 对 ab 个处理度量估计的交互作用; 存在 ab 个交互作用项
- MSAB 交互作用AB的均方: $MSAB = SSAB / [(a - 1)(b - 1)]$
- SSE 误差平方和
- χ^2_R Friedman的秩方差分析检验统计量

导 言

本书介绍的是最重要的近代统计方法，是帮助人们作出各种决策的统计学概要，只要求读者具备有限的数学知识就可以理解并利用本书的内容。实践证明：作者原来的德文版《应用统计》（第五版）对于科学研究工作者和咨询统计学工作者是十分重要的教科书，而现出版的本书的英文本又认真地做了校正和补充。

本书收集了各种应用统计方法，同时给出了这些方法在计量观测值及计数观测值的应用。本书避开了抽象的数学概念及其由来，把重点放在统计公式的基本原理以及对某个公式或某种检验有效的条件的解释上，优先考虑小样本和分布自由的方法。作为一本教科书和参考读物，本书对技术员，工程师，行政管理人员，大中专学生，物理学工作者，医学工作者，其它科学领域的研究人员，实际运用统计数学的工作者是一本易读书籍。

实际应用是本书的主导思想。本书的基本部分包含有 440 个已完全计算准确的数值例子。其中一些既实用而又非常简单明了。本书附有 57 个练习，收集了既有不同的计算方法，又有广泛的参考书目和非常详细的索引。特别本书收集了 232 个数学和数理统计表，使计算既简化而又能较顺利地进行。

现在简单地说明本书的结构：首先是关于某些初等数学的注意事项（中译本已将这部分删去——译者注），第一章是考虑统计决策技术，第二章是医学统计，序贯分析，生物测量，工业统计及运筹学等领域的介绍。数据样本和频数样本比较在第 3、4 章，以后三章涉及更进一步的方法：关联性分析（相关和回归，列联表分析）和方差分析。本书其它部分是一个综合性的一般的和专门的目录、练习、主题索引和作者索引（限于篇幅，中译本删去了作者索引和主题索引——译者注）。

标有箭头►的各节是最重要的统计技术的概述：1.1, 1.2.1—3, 1.2.5, 1.3.1—7, 1.4, 1.5, 1.6.1—2, 1.6.4—6, 3.1.1—2, 3.1.4, 3.2—3, 3.5, 3.6, 3.9.4—5, 4.1, 4.2.1—2, 4.3, 4.3.1—3, 4.5.1—3, 4.6.1, 4.6.7, 5.1—2, 5.3.1, 5.4.1—3, 5.4.5, 5.5.1, 5.5.3—4, 5.5.8—9, 5.8, 6.1.1, 6.1.4, 6.2.1, 6.2.5, 7.1, 7.2.3, 7.3—4, 7.6—7。

学习本书的比较非正式的方法是首先从头到尾阅读书内的材料，然后再读一遍统计学导论及以下各节：1.1, 1.2.1, 1.3.2—4, 1.3.6.2—3, 1.3.6.6, 1.3.8.3—4, 1.3.9, 1.4.1—8, 1.5, 3（序），3.1.1, 3.2, 3.6, 3.8, 3.9（序），3.9.4, 4.1, 4.2.1—2, 4.5.1, 4.6.1（止表83），5.1—2, 5.3（序及5.3.1），6.2.1（止(6.4)）7.1, 7.2.1, 7.3.1及7.7。

作者在安排文章的次序时发现某些困难（不能完全避免有少数例题要涉及以后几章的内容），也为了保持文章必须有的简捷性，作者建议初学者最好至少读两遍。只有那样，才能抓住各种相互关系以及理解统计知识最重要的先决条件。数值的例题，有些是非常简单的，能帮你加深对材料的理解及方法的应用，在某种意义下，这种例子相当于与数字打交道，且比真正来源于实际数据的例子（通常包括过多的数值计算）更有益，更有启发性，后者通常只为具有专门知识的专家所感兴趣。建议读者作为练习，独立地解出某些例子，并解决某些实际问题。

许多重要文献贯穿出现在正文中，它们指出了各种内在的相互关系，认真研究这些文献，有可能发现一些更进一步的结果。

我诚挚地感谢译者Z.Reynarowych先生，他的出色技巧使正文变得清晰易懂。

统 计 学 导 论

科学家和艺术家有理解外部世界的共同愿望，并希望减少世界明显的复杂性甚至混乱性，用一些有条理的表达式把它们表现出来。科学工作就包含着用有规则的方法表示无规则现象。

统计学是数据的艺术和科学：通过产生，收集，描述，分析，总结和表达数据达到发现新知识的目的。

统计学的基本任务：描述，确定，作出判断，得出与总体有关的推断。

每个人有时识别不出现存的关系和区别，而有时却凭空想象出并不存在的关系或分布。日常生活中，我们常常是凭借直观知识，即先天的理解力识别出相似性或差异性。科学家则发现某些新的现象、相关性、趋势或各种效果，基于这些，他们提出一个假设，然后，他必须检验此假设，判断所观测到的效果是否是在随机条件下产生的。至于判断观测到的现象是否可以认为严格随机或确实具有典型性，这个问题要由分析统计学来解决。分析统计学已成为近代科学方法的典型特征。

借助于统计方法，我们可以对一些**问题作出答复，考察一些陈述的有效性**。例如：希望对选举结果有一个大致的了解，谁会当选？每周两小时的体育锻炼对加强心脏功能和循环系统有益吗？哪一种牙膏可用于防止蛀牙？钢的质量如何依赖于它的成份？另一方面，我们可以观察到：售货员将每日营业额增加了\$1000；

通过治疗法A,某种疾病患者幸存的比例由60%提高到90%;化学肥料 K_1, K_2, K_3 对燕麦产量的效果不能分辨出大小;等等。

当所观测的数值具有代表性时,我们称之为数据,它们对于评价假设和发现新知识是非常重要的。

统计方法涉及从周围环境得来的数据及这些数据的收集和加工:即描述,估计,解释,**目的在于准备作出决策**。在十八世纪,“统计学”是用于鉴定各个民族条件的科学。数据是从整个人口,军事,工业及商业搜集而来的,因此导致了**描述统计**的发展。描述统计的任务是利用数据描述条件及事件,一般根据数据得出表格、图像、比率、指标和典型参数,如位置统计量(算术均值)和离散统计量(方差)。

分析统计可以从数据演绎出一般的定律来,其有效性超出了观测领域。它是从“政治算术”发展来的,“政治算术”的功能是根据洗礼、结婚及死亡登记来估计性别比例、人口出生、年龄结构及人口死亡的。分析统计又叫**数理统计**或**演绎统计**,其基础是**概率论**。概率论建立了包括随机性及随机试验在内的数学模型。下面是随机试验的例子:掷骰子,各种机遇和彩票的游戏,婴儿的性别,白天的温度,粮食产量,电灯泡的使用寿命,一次实验中测量仪器表盘的位置。一句话,凡结果受随机变量或测量误差影响的观察和试验都叫作随机试验。这里,我们**更感兴趣的不是数据本身,而是产生这些数据的原始总体**。例如:掷骰子,出现数字6的概率,或准备猜中一张彩票内6个数字的概率,或者1978年美国出生女性所占比例,这些比某些试验的结果更有趣。许多问题,包括可重复的试验,我们不能对试验和观测所有可能的结果(所谓的总体)进行观察,只能观察其中一部分。为了评价某种酒,我们只能从一桶酒中取出少量样本品尝,这个**样本**为我们提供了有关总体频数、成份及性质的信息,而整个总体或因基本的原因,或因成本、时间限制是不能全部研究的。