

图书馆统计学基础

[美] J. S. 辛普森 著

黄 巍 崔 磊 译

书目文献出版社

图书馆统计学基础

(英) I. S. 辛普森著

崔 巍 崔 岳译

书目文献出版社

tú shū guǎn tǒng jì xué jī chǔ

图 书 馆 统 计 学 基 础/(英)辛普森,I.S.著;
崔巍,崔岳译。—北京:书目文献出版社,1987.3

236页;32开

本书据原文《Basic statistics for librarians》2版译出
1.40元

I.图… I.①辛…②崔…③崔… I.图书馆统计

IV.G251.4 37.62 G25

图书馆统计学基础

(英) I. S. 辛普森著

崔 巍 崔 岳译

书目文献出版社

(北京文津街七号)

秦皇岛第二印刷厂排版

河北涿县辛庄印刷厂印刷

新华书店北京发行所发行 各地新华书店经售

787×1092毫米 32开本 7.875印张 130千字

1987年3月北京第1版 1987年3月北京第1次印刷

印数 1—12,900 册

统一书号: 7201·150

定价: 1.40元

内 容 简 介

本书是根据英国克莱夫·宾利公司1983年再版的《Basic Statistics for Librarians》一书译出的。该书自1975年出版以来,颇受读者欢迎,作者又根据大量的图书馆学教学经验,对其进行了修订。

这是一本为图书馆专业的学生和图书馆工作人员编写的统计方法入门读物,它所涉及的课题包括:数据资料的图示方法、平均数、离散、抽样、参数、统计量、概率、统计检验、方差分析、指数、时间序列和回归等。各种方法都用实例加以说明,这些例子取自统计学在图书馆工作和图书馆学中的实际运用。每章结尾都配有一组例题,用以检查读者对该章内容的理解。

前 言

图书馆专业的学生经常有这样一种思维逻辑：“统计同数字有关；数字是跟数学打交道的；我一向对付不了数学，所以我不能学统计学。”不错，学习统计学是要涉及一些数学的，但是，在初级水平上学习这门学问，需要的数学知识几乎不超过基本的算术规则。因此，在一般水平的数学课程中过不了关，并不说明统计学是难以理解的。事实上，以图书馆专业的学生——包括本科生、研究生和有过工作经历的人——为统计学教授对象的若干年教学经验已经证明，这门学科的基本内容是大多数人能够掌握的，不管他们以前的学习成绩或能力怎样。在更大的意义上，统计学是一种逻辑思维的学问，因此，学习美术的学生同数学家和科学家一样，也可以理解它——正如在计算机科学中，许多程序制作人员是艺术家出身，而不象人们经常认为的那样，是应用数学家。

当然，统计学的学习一直可以提高到攻取学位和更高的水平。在这些水平上，统计学的确是一门高度复杂的学问。然而，图书馆管理人员很少具备——或者被认为应该具备——如此高水平的专门知识。如果他们必须使用比较高深的统计技术，他们会去向一位胜任的统计学家请教——正如他们向设计师请教有关新图书馆的设计问题，或者向律师请教有关图书馆经营中的法律问题，或者向计算机专家请教如何实现图书馆管理的机械化。但是，在同专家谈话的时候，图书馆管理员应对谈论的问题有一定的了解，这是很重要的。因

此，这本书的目的就是使那些对统计学一无所知的图书馆管理员对这门知识、它的方法和运用有一个正确的认识。掌握了此书的全部内容，并不能使您成为一名训练有素的统计学家。

本书可用作大学图书馆专业学生的教科书，它能使学生对统计方法有足够的了解，从而使他们在图书馆管理和目录管理等专业领域里遇到问题时，知道如何使用这些统计方法。学生不必对有关的理论完全了解，但是对那些用来进行典型计算的经验方法是应该掌握的，可以用它们来解决其他类似问题。书中配有大量的示例。其中有些涉及实际的资料 and 情况，有些则不涉及实际问题，而使用假设的资料和情况，其目的仅仅是说明各种方法的运用。掌握这些方法，将有助于向学生表明：在实际工作中收集到的数据资料可以经过处理，得出非常有用的结论。这些方法也能帮助学生更有效地描述信息和说明结果。

从事实际工作的图书馆人员迄今一直是凭经验和直觉来解决工作中所遇到的问题的。然而他们正在发现，日益增加的压力和困难如今要求他们采用更加客观、现实的工作方法。他们也将发现，这本教科书将使他们获得统计学的基础知识，并在此基础上，通过学习较高水平的著作，确立他们对这门科学的信任感。

在本书的修订过程中，我采用了一些自本书第一版于一九七五年出版以来我在教授这门课程中使用过的内容。这些改进，是在同“纽卡斯尔图书与情报研究学院”的几位同事经过有益的讨论以后取得的。特别要提到的是罗杰·伍德豪斯和迈克尔·海因内两位先生。他们共同承担了“研究方法”

和“统计学”这两门课程的教学。同样，我非常高兴地接受了在我的指导下完成了这门课程学习的学生们所发表的见解。他们提出的意见和遇到的困难自然引起了我的注意，从而使未来的学生可以从他们的尝试和失败中获得教益。最后必须提到的是，我非常感谢我的妻子玛丽给予我的鼓励，没有她的鼓励，这本书的写成是不可能的。

I. S. 辛普森

于英国纽卡斯尔

一九八二年三月

绪 论

为了对统计学的学习有一个全面的认识，我们一开始就应对统计学这个术语的含意有一个尽可能确切的了解。统计学的定义不止一个。不过，《简明牛津英语辞典》所下的定义非常有用：统计学是“收集用数字表示的事实与资料，并对这些资料进行分类的学问。”这个定义使我们知道，资料就是“已知或假设的事实，是推理和计算的基础。”图书馆收集资料是为了保留不能靠记忆来保留的事实。这些资料通常出现在年度报告之中，如图书馆管理员人数、读者人数、馆藏图书的种类数、借阅图书的种类数等等。这些数字还经常同往年的可比数字一起出现。通常用来表现这些资料的方式能不能最充分地反映它们的特点？是不是最有助于读者理解它们？这仍然是一个有待争论的问题。用图表把信息传达给读者，也许比一长串数字更能说明问题，虽然这样会在出版时增加排版和绘图方面的费用。尽管如此，图表仍不失为表示数字事实的最重要的手段之一。但是，这种“描述统计学”在统计学这门学科中也许只是比较简单的部分。比较复杂的部分——根据已知的资料进行推理和运算——对于决策是十分有用的。七十年代和八十年代的经济困难和问题，使人们越来越清楚地认识到，图书与情报服务行业的管理必须依靠对有关因素的更加客观的估价，而不能仅仅依靠管理人员的主观经验和常识。某个管理人员也许会感到图书馆的读者在一定时期内比去年同期增加了。他也许会收集一些资料，看看

每天出借了多少图书，或者每周有多少读者走进图书馆。这种收集资料的工作虽然枯燥无味，但却是简单易行的。然而，单凭这样一些感觉和关于图书借阅量或读者人数方面的简单记录并不能使这位管理人员知道，这些每时、每天、每周、每年都在发生的变化对他是否具有重要的含义：他是否应该重新安排他的工作人员；是否需要增加人手；有无必要增加藏书量；应不应该调整借书卡与藏书量之间的比例；是否应在某些时间闭馆，等等。显然，如果收集的资料能够得到解释并被运用于管理决策，而不仅仅是以原始资料的形式存在，它们的价值就会大得多。

在这里，似乎有必要引用一句狄斯雷利说过的古老格言：“谎言，可恶的谎言，与统计同在。”用心不良的人当然可以通过统计资料的表现形式，通过对统计资料进行处理或解释，来愚弄粗心大意的人，但是这种现象并不应该归咎于统计本身，相反，这更加证明，图书馆管理人员应该熟悉各种收集资料、展示资料 and 解释资料的方法。

资料的收集方法对学习“研究方法”这门课的人具有更重要的意义。在这方面，读者可以阅读莱恩、弗德勒、布什和哈特的教科书。然而，读者必须明白，统计资料无论用多么高明的方法展示出来，也不如收集到高质量的资料本身。假如有一项调查是在某个下雨天，或者是在考试前夕，或者是在学校放假期间进行的，那么根据这项调查得出的有关图书使用情况的结论是毫无意义的。

另一方面，即使资料收集得很好，但如果对资料解释得很糟糕，或者结论只是从部分资料中得出的，那么统计学仍然很容易名誉扫地。假如连续数年的支出记录表明，某图书

管理员在购买图书时过于大手大脚，那么狠狠地把他训斥一顿并不难办到，但如果把支出记录同图书的成本指数比较一下，也许结果会表明，这个管理员应该受到表扬，因为他把支出的增长率控制在了通货膨胀率以下。统计学作为一种技术是令人满意的；值得怀疑的是这种技术的运用。

对于解释统计资料和滥用统计资料一类的问题，我们已经予以注意，因此，本书将不再对这些问题作进一步的论述。不过，希望读者在对本书论及的统计方法有了较多的了解以后，千万不要忽略这些方法存在的缺陷以及避免这些缺陷的方法。关于滥用统计资料问题的参考文献，可阅读坎贝尔、赫弗和金伯尔的有关著作。

由于统计学是与资料密切相关的，因此，读者应在一开始就懂得，资料是由各种具有不同属性的观察结果组成的。按照这个定义，图书馆的书架具有存放图书的属性，并且，如果我们逐个地观察每一个书架，我们就会发现，每个书架上的图书是不一样多的。某个书架上的书可能是二十本，另一个书架可能是二十四本，第三个书架也许是二十五本，等等。书架上的书的册数是一个变量。不过，变量可以分成两类：量别变量和质别变量。

量别变量是指每次观察都包含一次计数行为的变量。例如：一个书架的载书量就是一个量别变量，因为每观察一个书架，这个书架上的书的册数就要计数一遍。

同样，某图书馆的日图书借阅量也是一个量别变量，因为每观察一天，这一天借阅的图书的册数就要计数一遍。

并且，每次查找信息所占用的时间也是一个量别变量，因为每完成一次查寻，用来完成这项工作的分钟数就要计数

一遍。

相比之下，质别变量则是把某项观察结果按照规定的属性归入一个特定的组别或类别。例如，一本书的科目就是一个质别变量，因为任何一本书都可以归入一定的科类，取决于它是历史书、化学书、地理书，还是其它书。

同样，著作的类型也是一个质别变量，因为任何一件作品在登记入册时，都可能被归类为教科书，报告、专利说明书或期刊。

此外，出版物的语种也是一个质别变量，因为对每件出版物而言，它使用的语言可能是法语、德语，也可能是俄语或者汉语。

变量的类型必须能够简单而明确地辨别出来，这一点非常重要，因为资料的表现形式和资料的统计处理都取决于有关变量的类型。有些观察结果只是量别变量；有些则只是质别变量。但在某些情况下，观察结果可能既是质别变量，也是量别变量。而在另外一些情况下，同一观察结果可能会是两个不同的量别变量。本书对上述各种情况都作了论述，并且介绍了用图表形式表现统计资料的正确方法以及与之相关的统计运算。本书涉及的理论知识足以使读者对本书所论述的各种方法有一定的理解。关于基础知识，本书尽可能地减少了这方面的要求。如果读者感到需要复习一下必不可少的初等算术和计算方法，可以参阅附录一。

例题

下述变量是量别变量还是质别变量？

1. 某书的高度。
2. 某报告的寿命。
3. 某图书馆读者的性别。
4. 某图书馆读者的年龄。
5. 某平装书的成本。
6. 某读者从家到图书馆的距离。
7. 某文件的使用频率。
8. 某图书馆读者的居住地点。
9. 重新装订某书的必要性。
10. 某图书馆服务的人口数量。

目 录

前言	(I)
绪论	(1)
第一篇 量别资料	(1)
第一章 量别资料的表现形式	(1)
第二章 平均数	(16)
第三章 离散	(29)
第四章 抽样	(44)
第五章 统计量与参数	(57)
第六章 概率	(65)
第七章 统计检验	(88)
第二篇 质别资料	(113)
第八章 质别资料的表现形式	(113)
第九章 χ^2 检验	(129)
第三篇 质别/ 量别资料	(141)
第十章 质别/ 量别资料的表现形式	(141)
第十一章 指数	(151)
第十二章 时间序列	(163)
第十三章 方差分析	(175)
第四篇 量别/ 量别资料	(181)
第十四章 量别/量别资料的表现形式	(181)
第十五章 相关	(190)
第十六章 回归	(205)
例题答案	(213)

附录一	数学方法.....	(219)
附录二	常用符号.....	(225)
附录三	常用公式.....	(227)
附录四	随机数.....	(230)
附录五	t 分布.....	(231)
附录六	F 分布.....	(233)
附录七	X^2 分布.....	(235)
附录八	e^{-x}	(236)

第一篇 量别资料

第一章 量别资料的表现形式

根据资料的数量和性质，以及需要传达的信息的类型，表现量别资料的方法有以下几种。

线性排列

假如对一个变量所作的观察的次数不多，那么观察的结果可以记录成原始资料的形式。例如，某图书馆在十天中每天出借的图书数量分别为：

329 247 262 356 278 280 240 284 304 220

然而，如果按照数值的大小把这些数据由小到大或者由大到小排成一行，这个变量的值所包含的意义就会更容易理解：

220 240 247 262 278 280 284 304 329 356

如果用图形表示，这些值可以由图1.1中的一组呈线性排列的点来代表，因为每个变量的值只是一次观察的结果。

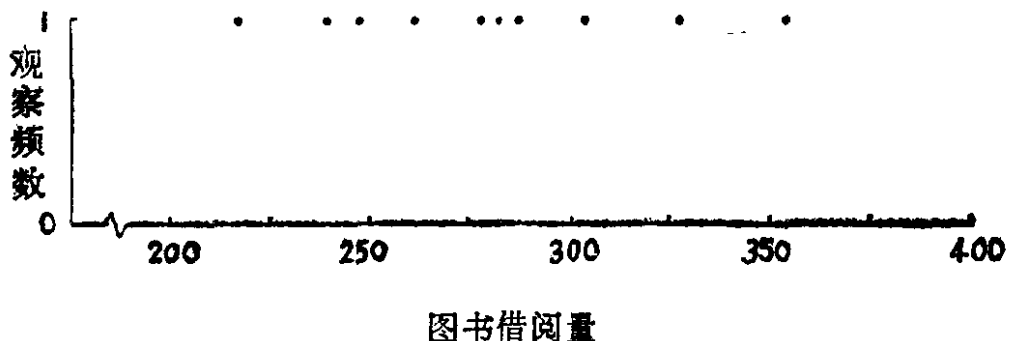


图1.1 十天中每天出借的图书数量示意图

如果对这个变量的某个特定值进行了两次观察，则代表

这个值的点的高度将是原有高度的两倍。

因此，从统计学的角度看，少量的观察是没有多大价值的，其唯一的作用就是表现变量的伸延范围。变量的最大值与最小值之差，称为全距。从图1.1的资料来看，全距为136，即：356—220。

频数表

随着观察次数的增多，原始资料的记录变得越来越难处理。比如，我们对某图书馆的八十个书架的载书量进行观察。按照观察的顺序，收集的资料如图1.2所示。

25	23	23	24	23	29	22	27	25	21
27	24	24	28	23	23	28	24	25	23
28	24	25	27	23	19	25	22	24	20
28	25	27	23	22	21	27	28	26	21
26	27	26	25	30	25	23	23	20	25
29	21	21	26	20	23	26	21	24	30
29	21	22	23	22	26	25	23	26	26
23	26	28	24	19	24	24	24	25	25

图1.2 书架载书量—原始资料

即使把这些资料按顺序排列出来，如图1.3所示，也没有多大用处。

然而，这些资料也可以用比较简明的方法来表现，这种方法可以使读者发现某种规律。图1.4把图1.3的资料用两行数字表现出来。上面一行是变量(每个书架上的图书数量)所有的值的排列；下面一行是变量的每个值被观察到的次数。后者称为频数，这种表现资料的形式称作频数表。

19	19	20	20	20	21	21	21	21	21
21	21	22	22	22	22	22	23	23	23
23	23	23	23	23	23	23	23	23	23
23	24	24	24	24	24	24	24	24	24
24	24	25	25	25	25	25	25	25	25
25	25	25	25	26	26	26	26	26	26
26	26	26	27	27	27	27	27	27	28
28	28	28	28	28	29	29	29	30	30

图1.3 书架载书量—顺序排列

书 架 载 书 量	19	20	21	22	23	24	25	26	27	28	29	30
观 察 频 数	2	3	7	5	14	11	12	9	6	6	3	2

图1.4 书架载书量—频数表

可以看出，自左至右依次递增的那行数字就是变量，而不规则地忽升忽降的那行数字是频数。因此可以说，我们观察到了两个载书量为19册的书架，观察到了三个载书量为20册的书架，等等。被观察到的书架的总数是所有频数之和，即：

$$2 + 3 + 7 + 5 + 14 + 11 + 12 + 9 + 6 + 6 + 3 + 2 = 80$$

相对频数

频数表的不足之处在于，我们无法一眼就看出载书量相同的那些书架在全部书架中占多大比例。这样，我们知道载书量为23册的书架有14个，但我们并不知道这个数字在所有的书架中是占一个较大的比例，还是一个较小的比例。为了