

809425

吉林省图书馆学会丛书之二十七

英国纽卡斯尔廷德图书情报学院

I·S·smpson 著

7541

7/1083

图书馆管理统计学基础

282
83

强 张宏钧 译
刘勇兵 校

教育部高等学校图书馆学
基本藏书

吉林省图书馆学会

7941
7/1083

000425

7941
7/1083

图书馆管理统计学基础

(英) I. S. Simpson 著

毕强 张宏钧 译

刘 勇 兵 校

吉林省图书馆学会

一九八二年·长春

内 容 简 介

本书是英国纽卡斯尔廷德图书情报学院 I.S.Simpson 为图书情报专业的本科生、研究生而设计编写的。本书介绍在图书馆工作中最常用的一些数理统计方法：定量数据的表达方式、平均数、离散趋势、抽样、典型统计量和参量、概率、统计检验、指数、时间序列、方差分析、相关关系和回归等。

本书对于图书馆工作如何运用数理统计方法作了扼要的论述，其内容充实，理论与实践并重。为帮助读者尽快掌握和运用，书中附有大量实例。

全书共有57幅插图，100多幅具体数据表格。

本书可作为图书情报专业的教材，也可作图书情报人员、应用数学工作者的参考资料。

Basic Statistics for Librarians

(英) I.S. Simpson 著

(第二版)

毕强 张宏钧 译

刘 勇 兵 校

*

吉林省图书馆学会出版 长春市新民大街10号

吉林省德惠县印刷厂印刷

*

开本：787×1092毫米1/32 字数：132 千字

1987年4月第一版 1987年4月第一次印刷

印数：1—1000册

吉印字：4028 定价：3.00元

前 言

图书馆专业的学生常常有这样一些想法：“统计学与数字相关；而数字又与数学有关；我过去从未运用过数学，那么我就无法研究统计学”。的确，统计学的研究涉及一些数学知识——但只涉及数学研究的基础，其要求不过稍高于懂得运算基本规则。所以，不能认为在数学方面稍差一些就无法弄懂统计学。事实上，多年来对图书馆专业的大学生、研究生及其毕业的后续课程的统计学教学证明，不论学生从前的学习和能力如何，大多数学生是可以掌握统计学的。统计学属于较高级别的逻辑思维，而不属于数学。正因如此，统计学不论对于学艺术的学生还是数学工作者或科技工作者都是同样可以理解的——这如同和许多学艺术的学生可以设计出很好的计算机程序而不必是应用数学家一样。

当然，学统计学可以达到获得学位或更高水平，达到这种水平，统计学是相当高深的。但是，很少有图书馆员具有——被要求掌握——这样一种高水平的专业知识。如果他们需要使用更高深统计技术的需要，那么他们可以请教一个合格的统计师——就如同他们要请教一个设计师有关一个新图书馆的设计问题，或请教一个律师有关图书馆管理出现的法律问题，或请教一个计算机专家有关图书馆事物处理机械化方面的问题。而且，在与专家的交流过程中，重要的是图书馆员应具备一些讨论该主题范围的知识。因此，本书的目的是为那些不懂统计学的图书馆员了解统计学、统计学方法及其应用而设计的。本书不适于专职统计员的培养。

该书是作为图书馆学专业学生的一本教材设计编写的。旨在使学生熟悉统计方法，使他们在研习某些专业课程如图书馆管理和书目计量学时，能够了解统计方法的应用。虽然学生们不可能达到对其基础理论的全面了解，但是书中给出典型计算的规则为学生提供了可以遵循和应用有关统计学方面问题的模式。书中多组实例，有些实例涉及真实的数据和情况，也有些是假定的数据和情况，这是为简单明了的说明各种统计方法的使用，而不是为了解决实际问题。掌握了这些内容，不仅会帮助学生描述他在研究计划中为得出有用的结论对收集的数据进行处理，而且还有助于学生更有效地提供信息和结果。

从事实际工作的图书管理人员直至今天在工作中仍然是凭经验和直觉，但是他们已感到日益增长的压力要求他们进行目的更为明确的研究。此教材可为这些人提供统计学的入门知识，并可据此参阅较专深的统计学著作。

在本书的编写中，我有机会又增补了我在自本书1975年第一版以来教学中使用的一些素材。这些素材是经与在纽卡斯尔廷德图书情报研究学院同行们的有益讨论，尤其是与同我讲授研究方法和统计学一样课程的罗杰·伍德浩斯和米切尔·海涅的共同讨论后增补的。机械化部分完全是基于有名望研究单位的经验和我在纽卡斯尔廷德工业计算中心的实践。最后需说明的是本书的编写得到了我夫人玛丽的大力鼓励。

I. S. Simpson

于 纽 卡 斯 尔 廷 德

1 9 8 2 · 5

吉林省图书馆学会会刊、丛书

编辑委员会

主 编： 金恩晖

副 主 编： 江乃武 曹殿举 赫彦生

编辑委员： （按姓氏笔划为序）

文甲龙 王一煦 江乃武

张凤桐 张 钧 邹秀玉

金恩晖 郭建文 曹殿举

符孝佐 赫彦生

目 录

前 言	(1)
导 论	(1)

第一篇 定量数据

第一章 定量数据的表达方式	(7)
第二章 平均数	(20)
第三章 离散趋势	(30)
第四章 抽样	(41)
第五章 典型统计量与参量	(52)
第六章 概率	(58)
第七章 统计检验	(76)

第二篇 定性数据

第八章 定性数据的表达方式	(97)
第九章 X^2 检 验	(112)

第三篇 定性/定量数据

第十章 定性/定量数据的表达方式	(125)
第十一章 指数	(134)
第十二章 时间系列	(144)
第十三章 方差分析	(155)

第四篇 定量/定量数据

第十四章 定量/定量数据表达方式·····	(163)
第十五章 相关关系·····	(172)
第十六章 回归·····	(185)

附 录

1 习题答案·····	(195)
2 使用符号·····	(198)
3 使用公式·····	(200)
4 随机数·····	(204)
5 t 分布·····	(205)
6 F 分布·····	(206)
7 X^2 分布 ·····	(207)
8 e^{-x} ·····	(208)
译者的话·····	(210)

导 论

要了解统计学，我们首先要尽可能地搞清楚什么是统计学。虽然统计学有几种定义，但是简明牛津英语词典下的定义较为有用，其定义是“统计学就是‘统计数字’或对数据的收集和分类”。何谓数据，数据就是那些所谓的“已知或假定的数字的实例，这些实例构成了推导和演算的基础。在图书馆中，要无休止地收集数据并将其通过年度报告进行表述，而表述又通常与前几年的数据进行比较，如工作人员数、读者数、贮存书刊数、流通书刊数等等。虽然图表表述法编制费时，又会为出版造成困难，是否是最好的数据表述方法，是否最有利还有待探讨，但是在向读者传递信息方面要比一连串的数字更有意义。图表是表达数据的一种重要手段，图书管理人员应该熟悉其编制和表达的方法。然而，这种“描写统计学”或许是统计学的简单部分，较难部分——根据数据进行推理或演算——才是决策的基础。本世纪七十和八十年代的经济压力使人们日益认识到图书馆管理和情报服务必须基于对相关因素更加客观地评价，而再不能只凭主观经验和馆员的一般常识了。他可以感到一周的某几天利用图书馆的读者人数多于其它几天；他可以感到在确定天数里利用图书馆的读者比前一年相应时间的人数多；他可搜集每天借出图书量的数据和每周进入图书馆的读者数。这种数据收集虽繁烦却极其容易，然而他的感觉和仅仅记录借阅量或读者数并不会告诉他年年、周周、天天、时时这些数字

是否有变化。以使他足以据此而调整人员或增加人员，调整他的图书贮量，更改读者借书卡的图书数，决定闭馆时间等等。虽然，如果对提供数据并用来做管理决策而不仅仅是描述其原始状态，那我们从中会获得更高价值。

迪斯雷利的演说词中有这样一句话，“统计数字是扯谎，弥天大谎。”当然，提供数据粗枝大叶，处理和描述数据的方法马马虎虎会蒙蔽粗心大意的人，但是这并不是统计学的过错。这更说明了图书馆员应该熟悉数据收集、描述和提供的方法的重要性。

数据的收集方法更接近于一种对研究方法的研究。关于研究方法读者可参阅莱茵等人编著的教材，可是读者必须清楚数据收集的质量比统计本身更为重要。在雨天对图书馆利用的调查或在大考前或假期对图书馆利用的调查对决策都毫无用处。

既使数据收集得好，但是如果其表述不当或基于部分数据就下结论，也会玷污其统计，而且很容易因为根据上一年的经费数据而使图书馆员大量浪费用于购置藏书资金。然而如果把这些数据与图书馆成本指数相比较，则可以告诫图书馆员避免资金的挥霍。统计技术是令人满意的，但如应用不当，也可出问题。

该教材提醒读者注意诸如数据的表述和滥用统计学的问题，但这并不是该教材的要旨，对此也没有做较深的探讨。尽管如此，我们希望随着读者对所描述的统计方法的了解，便可以避免上述的问题和错误。有关统计学滥用问题的可读性论述请参阅坎姻贝尔等人的著作。

因为统计学与数据相关，所以读者在开始就应懂得，数据是由一次观测变化到另一次观测的具有某些属性的观测所

组成。因此，一个图书馆中其书架具有存贮图书的属性。从一个书架到另一个书架，我们会发现书架上的图书数是变化的。一个书架可能有20本书，而另一个书架可能有24本书，第三个书架可能有25本书，等等。每个书架的图书数是一个变量。而且，变量具有两种类型，一种定量的变量和一种定性的变量。

一个定量的变量包括每次观测的统计量。例如，每个书架的图书数是一个定量的变量，因为要观测每个书架，统计的就是有关的图书数。

同理，一个图书馆日借阅是一个定量的变量，因为要对每天进行观测，统计的就是借阅的图书量。

加之，实现一次情报检索花费的时间是一个定量的变量，因为为实现每次检索，我们可以统计该工作花费多少分钟。

反之，一个定性的变量包含依据关注的属性对某一特殊组或类进行的一次观测，例如，因为观测的每册书可以认为是对历史书、化学书、地理书按类进行的观测，所以一册书的主题就是一个定性的变量。

同理，文献的类型是一个定性的变量，因为，由于对每篇文献进行了分类，所以这个变量可以被看成是一本教材、一份报告、一件专利说明书和一种期刊。

一个出版物的语种也是一种定性的变量，因为对每篇文献的查阅，可以发现是法文或德文或俄文或中文。

因为数据的表达方式和数据的统计处理取决于相关的变量类型，所以重要的是简单而又明确地识别变量的类型。有些观测仅含有定量的数据，有些观测仅含有定性的数据。而且在一些情况下，观测既包括定量的数据又包括定性的数

据。在另一些情况下，每次观测可以包括两种定量的变量。这两种情况在本教材中都予以了论述，并对用表和图的方式以及相关的统计运算表述数据的适当方法进行了解释说明。该教材还讨论了足以使读者理解所描述的统计方法的理论。考虑其实用知识，该教材尽可能不含假设。读者如需复习必要的基本运算方法和数学，请参阅附录。

习 题：

下述变量是定量的还是定性的？

- 1、一册书的厚度。
- 2、一份报告的年令。
- 3、一个图书馆用户的性别。
- 4、一个图书馆用户的年令。
- 5、一册平装本的成本。
- 6、从一个读者家到图书馆的距离。
- 7、一篇文献的使用频率。
- 8、一个图书馆用户的居住地点。
- 9、重新装订一册书的必要性。
- 10、一个图书馆服务的总体范围。

第一篇 定量数据

第一章 定量数据的表达方式

定量数据的表达方式可能有许多方法，其取决于所需数据的量和质以及信息的种类。

线性排列

如果观测的变量数量少，那么就可以把它们作为原始数据记录下来。例如，某一图书馆10天内图书借阅量分别为：

329 247 262 356 278 280 240 284 304 220

如果把上述数据按其量值的递增或递减顺序表示：220 240 247 262 278 280 284 304 329 356，我们就能够更易于理解这些变量的数值。

因为对变量的每个数值只观测一次，所以，我们可以用一些线性排列的点简单形象地描述它们（如图1.1所示）。

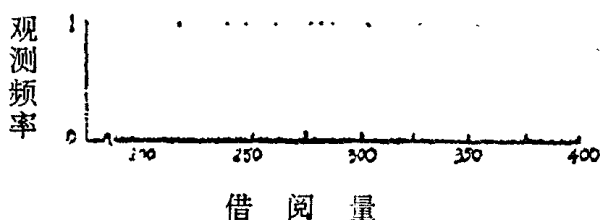


图1.1：10天内每天图书借阅量的示意图

若要对这些变量的某一特定值进行观测，圆点则将出现在二倍观测频率的高度上。

因此，从统计学的角度而论低频次的观测价值不大，其低频观测旨在说明变量的分布。变量的最高值和最低值之间的差别就是习惯上所说的排列。就图1.1中的数据来说，排列是136，即356—220。

频 率 表

随着观测频数的增加，原始数据的记录将变得越来越不容易处理。为此，把经观测的某图书馆中80个书架的图书数所得数据按顺序整理成表1.2的形式。

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

表1.2：每个书架的图书数—原始数据

如表1.3，尽管把上述数据加以排列也没有多大用处。

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

表1.3：每个书架上的图书数—排列

可是，我们能够用较为简洁的方法表示上述数据，并构成一种使读者一目了然的某种模式。表1.4是把表1.3中的数据成两条直线加以描述。上线是指排列中的变量值的集合（每个书架上的图书数）。下线是指变量的每个数值的观测次数。后者指的是频率，而整个表达形式就是通常所说的频率表。

每个书架的图书数	19	20	21	22	23	24	25	26	27	28	29	30
观 测 频 率	2	3	7	5	14	11	12	9	6	6	3	2

表1.4：每个书架的图书数—频率表

应注意，变量是指数字在排列中横向增长，而频率则是数字按纵向上升或下降。所以，观测两个书架，各有19本书；观测三个书架，各有20本书，如此类推。我们不难看出，观测的书架总数也就是所有频率的总和。

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

相对频率

频率表的局限性是不易看出给定书数的书架总数的确切比率。所以，我们可以看出，各有23本书的书架为14个，但问题是，书架总数的比率是大还是小呢？采用相对频率表来表述上述数据就能够解答这一问题。相对频率不是描述实际频率（如表1.4所示），而是通过实际频率的计算求得的，其表达式为：

$$\text{相对频率} = \frac{\text{实际频率}}{\text{所有频率的总和}}$$

因而，观测的各有23册书的书架的相对频率为 $\frac{14}{80}$ 或 0.175。计算后的全部数据如表1.5所示。现在我们能够直接