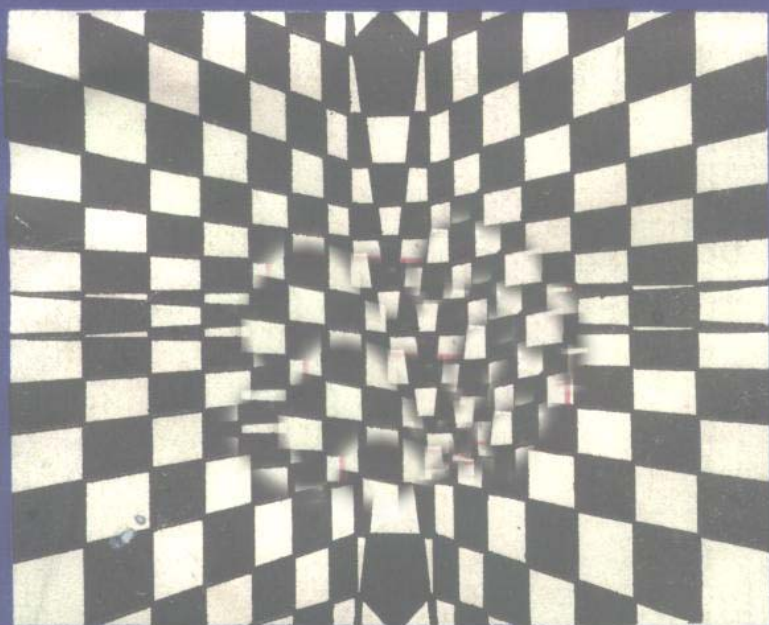


图书情报学 定量方法

- [印] 拉维昌德拉·劳 著
- 王崇德 黄 月 译
- 包昌火 胡新振 校



兵器工业出版社

G350
15

图书情报学定量方法

〔印〕拉维昌德拉·劳 著

王崇德 黄 月 译

包昌火 胡新振 校

DI59/10

兵器工业出版社

图书情报学定量方法

〔印〕拉维昌德拉·劳 著

王崇德 黄 月 译

包昌火 胡新振 校

*

兵器工业出版社 出版发行

（北京市海淀区车道沟10号）

新华书店总店北京科技发行所经销

北京海淀昊海印刷厂印装

*

开本：787×1092 1/32 印张：12 字数：258千字

1989年7月第1版 1989年7月第1次印装

印数：3000 定价：14.00元

ISBN 7-80038-048-3/G·9

内 容 简 介

《图书情报学定量方法》是由印度著名情报学家拉维昌德拉·劳编著、南开大学情报系和机电部兵器情报所翻译的一本定量方法在图书情报工作中应用的专著。全书分统计学方法和文献计量学方法两大部分，重点介绍了统计学基础、概率论、概率分布、相关分析、取样方法、假设检验、非参数检验、方差分析和文献情报流的基本规律以及它们的实用。内容翔实，文字通顺，虚实并举、实用性强。

作者简介

拉维昌德拉·劳 (Ravichandra Rao, I.K) 博士现任职于加尔哥答 (Calcutta) 印度统计研究院 (ISI) 班加劳尔 (Bangalore) 文献研究与培训中心 (DRTC)。

拉维昌德拉·劳博士在乌杜庇 (Udupi) 马哈他玛·甘地纪念学院获得数学与统计学理学学士学位后, 于1968年升入印度统计研究院攻读硕士学位。1970年他获得统计学硕士学位, 同时还获得了印度统计研究院的计算机科学学位证书。

拉维昌德拉·劳博士1970年获得硕士学位后即留校任教。此后, 他一直在文献研究与培训中心从事应用统计学、文献计量学、程序设计、图书馆自动化、数据库管理等学科的教学工作。他教数据库管理这门课程时特别着重于目录数据库的管理。

70年代中期, 他获得印度统计研究院文献研究与培训中心文献协会会员资格。1981年获得加拿大西安大略大学哲学博士学位。

1981年, 联合国教科文组织在马尼拉菲律宾大学图书馆学院举办了第4届情报科学研讨班。拉维昌德拉·劳应邀为学员们讲授了文献计量学和科学计量学这两门课程。

拉维昌德拉·劳博士在图书馆和情报科学领域的研究中有丰富经验, 造诣颇深。他的几篇学术研究论文很受称赞。

前 言

在过去的20年里，图书情报机构在馆藏建设和读者需求方面花费了相当多的经费，而这种势头至今犹未减弱。图书情报机构这样迅速发展的结果，引起了对其社会经济效益的一些研究和评论。直至最近，所发表的一些评论都是某些个人或某个组织的见解，应该说这种评论是主观的。然而，数学、运筹学、统计学、管理科学、经济学、系统分析等学科发展的结果，已将这种主观性减少到最低程度。特别是系统分析科学的发展，使得对图书情报人员所提供的情报服务的作用和效率有了合适的评价方法，这就是定量检验法。

但是，将定量检验法用于图书情报工作不是没有困难的。因为各个图书情报机构的服务种类和服务方式各不相同，所以要想建立一种能够适用于所有图书情报机构的通用的统计检验法，是十分困难的，或者说是是不可能的。下述三个方面使得这一问题更为复杂：

1. 各种图书情报机构及其服务的可资利用的数据各不相同，且不够用；
2. 情报激增（或称情报爆炸）；
3. 图书情报机构的用户日益增多。

主要由于上述这些问题，图书情报人员和读者经常发牢骚，他们抱怨，需要的情报在图书馆查不到，而查到的东西又不是需要的。为了解决这些问题，并控制日益增长的图书情报费用，图书情报人员开始在日常管理工作中应用定量检验法，特别是用这一方法来评价图书情报机构及其服务工作。

作。

图书情报人员就象产业界的经理一样，他们需要客观的数据，以便及时作出决定。产业界的经理所需要的经济数据是，产量与消耗、成本估算、销售收入、管理费用等经济计量学数据。对这些数据进行统计分析后，可以得到决策性情报，以便决定企业的采购项目、存货数量、快速推销项目、广告服务范围等。同样，图书情报人员也可利用下列数据作为文献计量的基本数据：参考统计、馆藏成本（价格分布）、图书馆性能指示图、图书采购及更换费用、流通数据、开馆时间、读者数量、脚注和引文、出版日期和地址、高等院校学术活动表、有关作者的出版物分布等。这些数据都可定量化，进行数学描述，这和处理经济计量学数据的方式是一样的。

本书的目的是，阐述统计学方法在图书情报科学中的应用，因此其讨论均以应用研究为基础。近年来，由于我们在这一领域的硕士生的教学特别强调统计学，所以希望本书能够成为学生、教师与图书情报人员的最有用的教材。1970年以来，作者即任教于班加劳尔文献研究与培训中心，本书即以作者在该中心文献与情报科学高级学习班教学时所用讲义为基础而写成的。

拉维昌德拉·劳 1983年7月
于班加劳尔

目 录

作者简介.....	(1)
前言.....	(2)

第一篇 统计学方法

第一章 引言	(1)
1.1 总论.....	(1)
1.2 图书情报学的研究性质.....	(2)
1.3 图书情报学研究的各个阶段.....	(3)
1.4 测量的级别.....	(4)
1.5 小结.....	(9)
1.6 练习.....	(10)
第二章 统计学基础	(11)
2.1 数据的表示.....	(16)
2.2 集中趋势的测度.....	(38)
2.3 离差的测度.....	(56)
2.4 小结.....	(69)
2.5 练习.....	(70)
第三章 概率论	(74)
3.1 定义.....	(75)
3.2 集合及其运算.....	(78)
3.3 概率.....	(79)
3.4 信息论.....	(85)

3.5	期望·····	(88)
3.6	小结·····	(90)
3.7	练习·····	(93)
第四章	概率分布·····	(96)
4.1	离散分布·····	(98)
4.2	连续分布·····	(110)
4.3	概率分布的参数估计·····	(119)
4.4	概率和理论频次的计算·····	(124)
4.5	变量的转换·····	(126)
4.6	举例·····	(128)
4.7	小结·····	(140)
4.8	练习·····	(141)
第五章	相关分析·····	(145)
5.1	散点图·····	(146)
5.2	相关系数·····	(147)
5.3	最小二乘法的曲线拟合·····	(154)
5.4	判定系数·····	(165)
5.5	小结·····	(166)
5.6	练习·····	(168)
第六章	取样方法·····	(171)
6.1	样本大小·····	(173)
6.2	取样方法·····	(175)
6.3	举例·····	(178)
6.4	小结·····	(180)
6.5	练习·····	(182)
第七章	假设检验·····	(184)

7.1	显著性检验	(184)
7.2	显著性检验: σ 是未知的	(190)
7.3	比率检验	(194)
7.4	显著性检验: 标准差	(198)
7.5	显著性检验: 双样本检验	(200)
7.6	关于总体相关系数的检验	(209)
7.7	对有关回归问题的假设检验	(213)
7.8	评估	(216)
7.9	小结	(221)
7.10	练习	(222)
第八章	非参数检验	(225)
8.0	引言	(225)
8.1	独立性的 χ^2 检验	(226)
8.2	连带测度	(234)
8.3	多项假设检验	(236)
8.4	总体的概率分布检验	(238)
8.5	科洛莫高洛夫-斯米尔诺夫 检 验	(243)
8.6	等级相关系数检验	(245)
8.7	其它检验	(247)
8.8	小结	(252)
8.9	练习	(252)
第九章	方差分析	(253)
9.1	单向方差分析	(253)
9.2	小结	(264)
9.3	练习	(265)
	参考文献	(266)

第二篇 文献计量学

1. 文献计量学的起源.....	(269)
2. 文献计量学的各种分布.....	(270)
2.1 齐普夫定律.....	(271)
2.2 洛特卡定律.....	(274)
2.3 分散定律: 布拉德福定律.....	(280)
3. 文献计量学分布特性.....	(295)
4. 等级与大小频次分布.....	(298)
5. 其它的文献计量学研究.....	(200)
5.1 引文分析.....	(200)
5.2 图书馆利用研究.....	(303)
6. 结论.....	(322)
参考文献.....	(326)
附录.....	(337)

第一篇 统计学方法

第一章 引言

1.1 总论

图书情报学的研究正在经历着重大变化。发展趋势表明，人们日益重视对图书情报机构日常工作进行逻辑的、科学的研究和定量化分析。趋势之一就是，建立各种模型，以描述图书情报机构利用的各个方面。图书情报机构的调查研究，将逐渐以分析的方法来取代简单描述的方法。分析模型可以帮助研究人员掌握图书情报机构的全面真实情况。模型的设计和发展是一个连续过程，即，根据现有可利用的数据和信息，对一些理论和观念进行试验、改进和发展。这样的科学研究工作主要包括以下几个方面：

1. 提出假设；
2. 实验的设计；
3. 观测与数据收集；
4. 数据分析——数据的分类与归纳，以文字、表格和图形表示的数据；
5. 模型的检验与发展；
6. 数据的分析与归纳。

需要时, 社会科学研究人员可以得到职业统计人员所提供的意见和报道。但是, 由于受到知识领域的限制, 这些人所提供的意见和报道往往是一般性的。从另一方面来说, 如果研究人员熟悉统计学和其他量化方法, 则他不仅能够理解统计人员所提供的意见和报道, 而且能够比较详细、深入地探讨这些数据 and 资料。因此, 研究人员如果具有定量分析的基本方法和知识, 则大有好处。

从这一点出发, 作者写了这本专著, 相信它对图书情报行业的学生和研究人员是会有帮助的。作者从 1970 年就在班加劳尔文献研究与培训中心任教, 本书是在作者为该中心文献与情报科学高级学习班所写讲义的基础上写成的。

1.2 图书情报学的研究性质

对于图书情报学的研究性质, 我们之中多数人都有这样的看法, 即: 形成假设, 并经常通过系统研究来验证这些假设。例如, 弗希尔(Fussler)和西蒙(Simon)于 1969 年曾这样假设: 一种文献的利用, 会随其年限的增加而递减^[1]。与此相反, 拉维昌德拉·劳则于 1981 年假设: 经常被借阅的文献比不经常被借阅的文献更容易被再次借阅^[2]。这些假设往往是根据过去的经验、直觉、典型研究等作出的。

与物理科学研究人员一样, 图书情报科学人员也经常进行研究, 以便对本领域的问题和争论有更深入的了解。图书情报学的研究有多种形式和方法, 研究的问题也很广泛。为了检验图书馆和情报中心的收藏是否充足, 借阅政策是否合理, 特定图书馆和情报中心计算机服务的效果如何等等,

研究人员可以开展读者调查、读者或文献取样调查、非公开出版物价值分析等方面的调查研究。

在必要的假设形成，并收集到有关数据以后，须做系统的检验，以确定这些假设是否正确。即或这些假设是合乎逻辑的，甚至是不言而喻的，也须做系统的检验。例如，麦克格拉斯 (McGrath) 于 1978 年就对其研究的下述假设作了验证^[3]：

1. 越是软科学性课题，需要的参考书越多，越是硬科学性课题，需要的参考书越少；
2. 理论性越强的课题，需要的参考书越多，应用性越强的课题，需要的参考书越少；
3. 时效性较强的课题与无时效性课题相比，所需参考书数量大不相同。

1.3 图书情报学研究的各个阶段

利用科学的方法来检验关于对某一图书情报机构的特点和情况的假设，需要进行周密的研究和计划。这种研究分下列几个阶段：

1. 确定研究课题；
2. 根据对统计资料的分析形成假设，起始过程的检验称为零假设，例如，借阅流通与馆内阅读之间没有关系；
3. 制定合适的研究设计；
4. 收集数据（通过记录性资料、调查表以及带着问题或不带问题的访问来进行，或通过机电设备或人工观测来进行）；

5. 数据的分析;
6. 结果的解释与总结。

本书以后的各章节与研究的数据分析阶段紧密结合。为了检验零假设,在这一研究阶段,将进行数据列表、计算、计数、总结、重新排序和比较。

1.4 测量的级别

为了提高研究质量,很多研究人员在进行数据分析时常采用量度或系列数字。对于图书情报研究人员来说,根据其采用的特殊测量级别,数字至少有三个重要作用。系列数字可用于:

1. 名称和计数——测量的名义级;
2. 等级排列——测量的顺序级;
3. 记分或标记——测量的区间级。

1.4.1 测量的名义级

测量的名义级是一种方法,用来命名或标志一个观测或一组观测。人们可以用名义测量来表示读者与流通文献之间的关系。表 1.1 是名义级数据的实例。

文献可以根据其利用频次加以分类。在进行这种分类时,可根据文献被借阅的频次将其分为常用和不常用两类。同样,读者也可分为常借阅和不常借阅两类。表 1.2 是另一个名义级数据的实例,根据文献被借阅次数和读者借阅特定文献的次数,将表中登记数字分为两组。

表 1.1 文献和读者的分布流通数据分析(测量名义级)

文 献 \ 读 者	大学在校生	毕 业 生	研 究 生	职 工	教 师	合计
图 书	70	75	50	90	30	315
报 告	10	25	70	10	98	213
标 准	3	10	23	2	39	82
专 利	2	5	31	1	28	67
期 刊	8	40	92	15	141	196
其 他	2	3	10	5	32	52
合 计	95	158	281	123	368	1025*

*系某图书馆在某一给定时间内登记的总数

表 1.1 和表 1.2 所列数据常常被称为复变名义数据。表 1.2 也称为 2×2 联列表。杜威编制的文献十进位分类表是名义测量级的另一个实例。按读者性别分组也是一种名义级测量。这样的分类只是任何名义级测量的初级阶段。反之,研究者的最高目标应该是,寻求并发现各种类别之间及其内部的各种因素(例如,公共图书馆读者内部的划分)。

研究人员应设法回答下列问题:

1. 图书馆读者中有社会背景的人**是否比非图书馆读者中有社会背景的人要多些呢?
2. 在公共图书馆的读者中大学生是否多于中学生呢?

**印度用来区分人们社会身份和地位的用语。

表 1.2 读者(以借阅频次为基础)与
文献的分布*(测量的名义级)

	借 出 的 文 献		
	一 次	一次以上	共 计
一 次	934	170	1,104
一次以上	16,277	3,898	20,175
共 计	17,211	4,068	21,279

• 摘自拉维昌德拉·劳的《Documents and user distribution in transaction records of Canadian University Libraries》一文(1981)。

为了回答这些问题,须在名义级层次上收集数据。名义级数据不分等级,不排顺序,也不按性质分类,例如,好或坏,高或低,等等。对图书馆读者的名义测量,如按其学术地位,则不可能清楚地表明研究生是否优于或低于本科生。名义级数据只根据名称或数字来进行标志,但是根据给定的性质和特征,可以将情况或观察分成独立的类别,以表明其相同或相异。

1.4.2 测量的顺序级

当研究人员进行名义级测量范围之外的研究,并试图按照研究对象具有的任何给定特征的等级将其观察进行排序时,他就按测量的顺序级进行工作。顺序类别之间相互关系