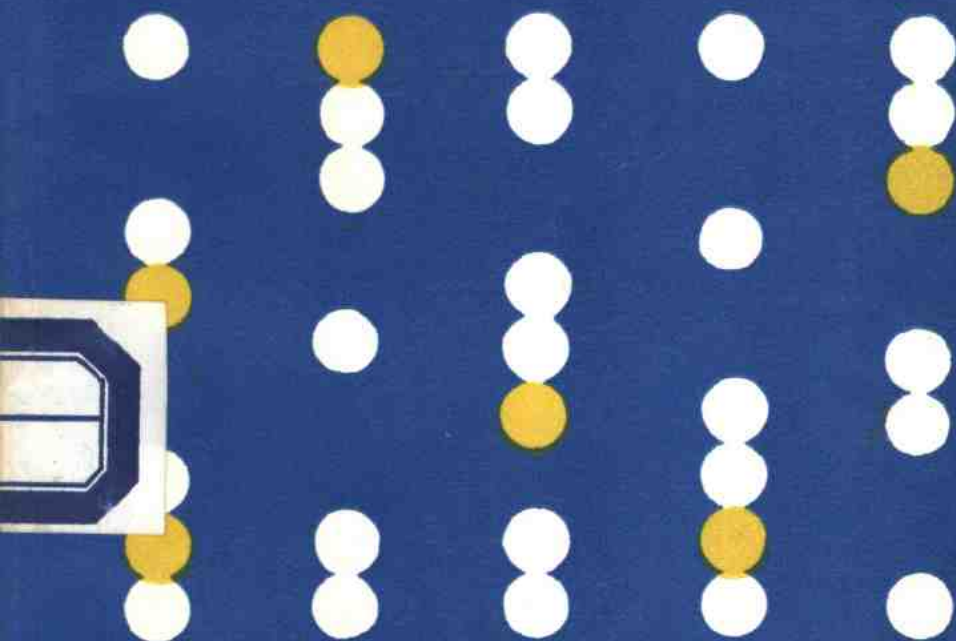


[日]大泽清二 著
季成叶 译

生活统计

de

基础知识



北京医科大学中国协和医科大学联合出版社

生活统计的基础知识

(日) 大泽清二 著

季成叶 译

北京医科大学
中国协和医科大学 联合出版社

085212

(京)新登字147号

日本国 家政教育社 出版

ISBN4-7608-0255-0 信每书籍印刷株式会社印刷

1990年5月10日 第1版第一次印刷

1992年4月20日 第1版第三次印刷

生活统计的基础知识

(日) 大泽清二 著

季成叶 译

暴海燕 责任编辑

* * *

北京医科大学 联合出版社出版
中国协和医科大学

(社址: 北京医科大学院内)

新华书店总店科技发行所发行 各地新华书店经销

北京医科大学印刷厂 印刷

* * *

开本: 850×1168 1/32 印张: 6.3125 字数: 164 千字

1993年5月第1版 1993年5月第1次印刷 印数: 1-2000册

ISBN 7-81034-221-5/R·221 定价: 7.65 元

内 容 简 介

本书以通俗的语言、简练的笔法从频数分布、平均偏差、正态分布、相关回归、假设检验到根据最新研究成果撰写的时系列资料及趋势分析等全面系统地介绍了统计学的基础知识。本书所举实例与现实生活十分贴近。如：运用统计图表帮助孩子制订复习计划，运用概率推算提高高考的命中率、利用统计信息扩大市场销售量……。本书把抽象的推理分析过程化解成生动浅显的生活语言，使读者在轻松的气氛中学到各种现代生活所必需的统计知识，并能在日常生活中得以应用。

著者略历

大洋清二（おおくさね せひし）

1946年生于日本国埼玉县

毕业于东京大学教育学系，获教育学博士

曾受聘为日本东南亚医疗情报中心研究员、日本国立筑波大学讲师

现任（东京）大妻女子大学人间生活科学研究所教授

译者略历

季成叶

1946年生于中国上海市

毕业于北京医科大学公共卫生学院，获医学硕士

1987～1989年期间公费留学日本，为日本国立筑波大学外国人研究员

现任 北京医科大学公共卫生学院副教授，日本国大妻女子大学人间生活科学研究所特别研究员

作者前言

20世纪还剩下最后10年。20世纪将以它取得的划时代进步，在人类历史上书写下光辉的一页。特别值得一提的，是电子计算机技术的诞生及其飞速普及，完全改变了我们的生活。

“现代”是什么？一言以蔽之，现代就是科技化情报化的时代。

生活在这个时代的我们，不能不掌握统计学知识，因为它是现代人类不可缺少的生活技能，也是人类智慧的结晶。然而遗憾的是，迄今为止不了解统计的重要性，或出人意料地对它表现出轻视态度的还大有人在。另一方面，不少人心里想学，但又觉得统计太抽象、太难理解，因而望尘却步。即便是大学里的教科书，问题也不少。有的只讲理论，不与生活实际相联系；有的列举了一大堆的公式推导，举了许多脱离实际的例子，却没有把基本的概念系统解释清楚，统计的推理和计算步骤也没能用浅显易懂的生活语言加以说明。这样教出来的学生只会考试，只会套用统计术语，却不会把它活用于生活实际。

作者担任大学的统计教师已有十多年，虽不敢说已完全精通统计这门专业知识，但对大学里、社会上存在的上述种种问题感知颇深，一心想为改善这一现状出点力。构思已久，只因日日穷于应付日常工作，没有时间下笔。正巧这次去泰国进行第七次泰国东北部地区学生健康和生活制度调查前因病住院，就在病榻上完成了本书的初稿。我在写作中力求语言通俗易懂，所举实例贴近现实生活。所以，这本书既可供个人自修，或作为一般大学的统计基础教材，也能为正在工作的人们解决实际生活中的统计问题所用。如果读者们能在较轻松的气氛中读完本书，从此消除对

统计学的畏惧感，并在日常生活中自觉或不自觉地用到书中的一些原理和方法，将是我最大的心愿。

大泽清二

写于泰国东北部地区第七次学生
健康和生活方式调查前夕

译者前言

自从十九世纪伟大的数学家克托莱、彼尔森、高尔顿、威尔登、艾德沃夫等人开创了现代统计学的先河以来，统计学的命运始终和工业化、近代化的命运紧紧相连。到了本世纪中叶，随着电子计算机技术的诞生和发展，各种现代统计技术也应运而生，成为科技进步的重要助推剂之一。然而，统计技术的发展又给其自身带来了难题：它那抽象的概念、繁琐的推理和枯燥的数字使许多人望而生畏，想学又无处下手；一些统计专著充斥着艰涩的术语、复杂的求证公式，却日益脱离统计学赖以生存的土地——现实生活。这无疑加深了亿万群众掌握现代统计技术的难度，也给电子计算机迈入千家万户带来了阻力。

日本著名社会统计学家大泽清二教授所著的《生活统计的基础知识》一书给人们带来了一股清新之风。该书共19章，从结构上可分两大部分。第一部分自第1至第11章，以通俗的语言、洗练的笔法，从统计图表、统计资料的衡量尺度、频数分布、平均值和各种离散度指标、正态分布和偏差值、相关分析和回归预测等，直到根据较新研究成果等撰写的时系列资料及趋势分析等，全面系统地介绍了统计学基础知识及其实际应用范围。第二部分自第12章到第19章，着重介绍抽样方法、区间推定、检验假设、相关系数检定，以及针对不同资料而采取的各种平均值、方差及标准差的显著性检验方法，等等。本书例举的各种实例和复习思考题也别具匠心，十分贴近生活实际。例如，怎样运用统计图表的制作原理帮助孩子制订复习计划；怎样通过概率推算，正确估计自己的实力，提高报考志愿的命中率；怎样利用各种统计信息准确地预测市场趋势，扩大商品销售量；以及如何根据本职工作的需要，设

计并开展一些统计调查，等等。著者还根据自己的多年教学实践，例举一些初学者容易混淆的概念和不易觉察的错误，和读者一起讨论分析，以加深理解。书中还配有百余幅图表，图文并茂，抽象艰涩的推理过程化解成生动浅显的生活语言，使读者在轻松的气氛中读完本书，学到各种现代生活必备的统计知识，并能见诸实际应用。

大泽清二先生是我留学日本期间的指导教师之一，又是科研工作的长期合作者。他崇尚实际、摒弃虚华、治学严谨的科学作风，在本书中随处留下印痕。大泽先生对辉煌灿烂的中华民族历史文化推崇备至，对中国人民怀有深厚的友情。当他得知我准备把本书译成中文出版时，不止一次地流露出由衷的喜悦，并一再委托我向他的中国读者们致意。今天，当我最后完成本书的译稿时，我的唯一遗憾是，由于本人水平有限，译文中难免有种种不妥之处，还望各位读者多多指正。

译者

1992年10月于北京

目 录

第 1 章 统计图表的正确使用	
—从有田健二的复习计划谈起—	(1)
第 2 章 统计资料的基础知识	
—尺度是什么—	(6)
1. 统计学的大数量法则	(6)
2. 统计资料的特征	(6)
3. 变量和原始资料	(7)
4. 计量变量和计数变量	(7)
5. 四种尺度标准	(8)
(1) 名义尺度	(9)
(2) 顺序尺度	(9)
(3) 间隔尺度	(11)
(4) 比例尺度	(12)
第 3 章 频数分布	
—使资料井然有序—	(14)
1. 频数分布	(14)
2. 研究频数分布的重要性	(16)
3. 资料的中心化倾向	(17)
第 4 章 最频值、中位数值和平均值	
—测定资料的分布位置—	(20)
1. 最频值	(20)
2. 中位数值	(22)
3. 平均值	(23)
第 5 章 信息量、分散系数和百分位数	
—了解资料的分散状况(I)—	(29)

1. 了解资料分散状况的重要性	(29)
2. 各种信息量	(31)
3. 范围和分散系数	(36)
4. 百分位数	(37)
第6章 平均偏差、方差和标准差	
—了解资料的分散状况(Ⅱ)—	(43)
1. 平均偏差	(43)
2. 方差	(44)
3. 标准差	(45)
4. 标准差计算中容易发生的错误	(46)
5. 变异系数	(49)
第7章 正态分布和离差值	
1. 正态分布	(52)
2. 正态分布表	(54)
3. 正态分布规律的发现	(56)
4. 离差值	(59)
第8章 相关系数和顺位相关系数	
—测定变量间相互关系的强弱—	(64)
1. 相关图	(64)
2. 相关系数的计算	(67)
3. 顺位相关系数	(73)
4. 相关系数大小的意义	(75)
5. 相关分析技术的由来和发展	(76)
第9章 回归分析	
—重要的预测手段之一—	(78)
1. 最小两乘法	(78)
2. 回归式的说明率	(84)
3. 从回归到相关系数	(85)
4. 回归分析技术的起源和发展	(86)

第10章 时系列资料分析	(88)
1. 移动平均	(88)
2. 移动平均中的加权	(89)
第11章 时系列资料的趋势分析	(94)
1. 趋势分析	(94)
2. 相对预测比	(98)
3. 对时系列资料中各变动因素的分析	(99)
(1) 倾向变动因素	(99)
(2) 季节变动因素	(99)
(3) 循环变动因素	(99)
(4) 不规则变动因素	(99)
第12章 统计调查的方法	(105)
1. 典型调查法	(105)
2. 随机抽样调查	(106)
(1) 单纯随机抽样	(107)
(2) 系统随机抽样	(109)
(3) 分层随机抽样	(110)
(4) 多段随机抽样	(112)
第13章 统计的推理	
——介绍区间推定的方法——	(114)
1. 正态分布与区间推定	(114)
2. 样本的扩大与区间推定	(115)
3. 未知总体标准差时的区间推定	(118)
4. t 分布的导入	(119)
5. 利用卡方 (χ^2) 作区间推定	(123)
第14章 检验假设的方法	(127)
1. 显著性	(127)
2. 无效假设和对立假设	(128)
3. 检验假设的基本程序	(128)

4. 双侧检验和单侧检验	(129)
5. 检验假设结果的误判	(130)
6. “显著性”和“显著性水平”	(131)
7. 显著性水平多大才适合	(132)
第15章 F 分布	
—比较资料分散状况的方法—	(134)
1. F 分布表	(134)
2. F 分布表的实际应用和注意事项	(136)
第16章 平均值差的检验	(139)
1. 比较平均值的各种情况	(139)
2. 第Ⅰ类的求解方法——z 检验(1)	(141)
3. 第Ⅱ类的求解方法——t 检验(1)	(142)
4. 第Ⅲ类的求解方法——z 检验(2)	(143)
5. 第Ⅳ类的求解方法——z 检验(3)	(144)
6. 第Ⅴ类的求解方法——t 检验(2)	(145)
7. 第Ⅵ类的求解方法——t 检验(3)	(146)
8. 第Ⅶ类的求解方法——t 检验(4)	(148)
9. t 检验前注意事项	(149)
第17章 相关系数的检验	(152)
1. 总体有没有相关	(152)
2. 总体相关是高还是低	(153)
3. 根据样本相关系数推定总体相关系数	(154)
4. 两两指标的相关系数有差别吗?	(156)
第18章 比率	(159)
1. 各种比率关系	(159)
2. 根据样本比率推算总体比率	(160)
3. 比较样本比率和总体比率	(161)
4. 比较两样本的比率	(162)
5. 利用交叉集计表进行分析	(163)

6. 价格指数.....	(168)
7. 消费者物价指数.....	(170)
第19章 充分利用统计资料.....	(173)
1. 官方统计.....	(173)
2. 目录和索引.....	(174)
3. 年鉴和白皮书.....	(175)
附表	
附表 1 乱数表.....	(178)
附表 2 正态分布表.....	(179)
附表 3 t 分布表.....	(180)
附表 4 χ^2 分布表.....	(181)
附表 5 F 分布表.....	(182)
附表 6 相关系数的显著性水平.....	(188)
附表 7 顺位相关系数的显著性水平.....	(188)
附表 8 自然对数表.....	(188)
附表 9 常用希腊字母对照表.....	(189)

第1章 统计图表的正确使用

——从有田健二的复习计划谈起——

现代社会是信息化的社会，现代生活离不开统计。不但你的学业和职业需要大量统计学知识，就是在你的日常生活中，也时时处处会碰到各种需要利用统计来解决的实际问题。学习统计，最好先从学会整理资料入手。整理资料，实际上就是对你已经获得的大量表面上看来杂乱无章的统计数字，通过深入观察，从中挑选出那些对你最有用的部分加以归纳整理的过程。统计图表，则是这一过程的具体形象体现。好的统计图表，会向你提供大量有意义的知识、信息和解决实际问题的线索。

在学习使用统计图表之前，不妨让我们先看下面这个实际例子。

有田健二是个小学六年级学生，立志报考东京都内某所名牌中学。眼看升学考试的日期一天天临近。可是复习的内容又多又乱，真不知该从哪儿下手。为此，他请教了正上大学的哥哥有田健一郎，请他帮助自己制订一个复习计划。

健一郎制订的复习计划先从理科入手。他把收集到的那所名牌中学最近10年来的理科升学考试试题，按教科书的内容归成类，标出每一章的平均出题数；然后又翻阅了健二最近半年内在学校里参加的分单元模拟考试的成绩单，随即画出下面这张表（见表1-1）。

把健二的模拟考试成绩和学校十年间的出题频度汇合在一起，根据下列公式可得到：

必需努力的程度(x) = (100 - 模拟考试得分) × 出题频度

公式右边的两个因素，前一个反映出健二在理科学习上的差

表1-1

健二君的理科重点复习时间分配

单元内容	模拟考分	报考学校 出题频度 (出题数)	(频数)	模拟考试 丢失分数	必需努力 程度 (%)	学习时 间分配 (%)
1. 人的身体	59	3	(4.6%)	41	188.8	5.5
2. 力的作用	85	4	(6.2%)	15	93.0	2.7
3. 物质的变化	95	3	(4.6%)	5	23.0	0.7
4. 天体	80	6	(9.2%)	20	184.0	5.4
5. 电流和磁石	77	12	(18.5%)	23	425.6	12.4
6. 物体的熔化	46	14	(21.5%)	54	1161.0	33.8
7. 水溶液性质	41	11	(18.9%)	59	997.1	29.0
8. 光学知识	72	6	(9.2%)	28	257.6	7.5
9. 生物和生长	86	4	(6.2%)	14	86.8	2.5
10. 岩石和土地变化	93	2	(3.1%)	7	21.7	0.6
合计		65	(100%)		3438.3	100.0

距, 后一个提示出题的可能性。健二的复习时间已经很有限了, 所以他首先要抓好自己的薄弱环节。学校的出题倾向在相当程度上反映出对某一方面知识的重视程度, 所以也应该是复习的重点。很明显, 表1-1的10个单元中, 哪个单元所得的 x 值越大, 健二在这方面应作的努力应越大, 投入的复习时间相应也就越多。

健二还告诉哥哥, 学校的老师提醒过他, “人的身体”、“水溶液的性质”、“生物与生长”和“岩石和土地的变化”这四个单元很重要, 要多花些时间复习。可是健一郎仔细看看他编制的表1-1, 摇摇头表示不同意。他的理由是: “人的身体”和“水溶液的性质”这两部分虽然要认真下功夫复习, 但从表1-1对复习时间的分配来看, 对此已作了充分考虑, 不需要再挤占更多的时间。至于“生物和生长”、“岩石和土地的变化”这两部分, 出题频度低、健二的掌握程度也很好, 即便少花点时间, 问题也不大。

上面这个例子, 就是对统计表的具体应用。当然, 实际生活

中的现象可能远比此复杂，不能预料的事情也很多。但只要能在充分掌握基本资料的基础上，合理运用你所掌握的统计知识，你就会比以往更准确地把握工作、学习重点，取得好成绩的可能性也更大。

根据健一郎制成的表 1-1，你还能按不同要求画出各种统计图。统计图是对统计资料的形象化描述，它通俗易懂，便于比较，而且直观。

图 1-1 是最简单的直方图。健二的各单元考试成绩是相互独立的，所以其数值能通过直条的长度表示。图上清晰显示出，在“物质的变化”和“岩石与土地的变化”等方面健二的成绩最好，但在“物体的熔化”和“水溶液性质”方面成绩最差，应加紧努力。绘制直方图特别要注意各单元的单位必须是一致的，而且横轴的尺度必须从 0 开始，否则直条长度的比较就不能真实反映数值之间的比例。

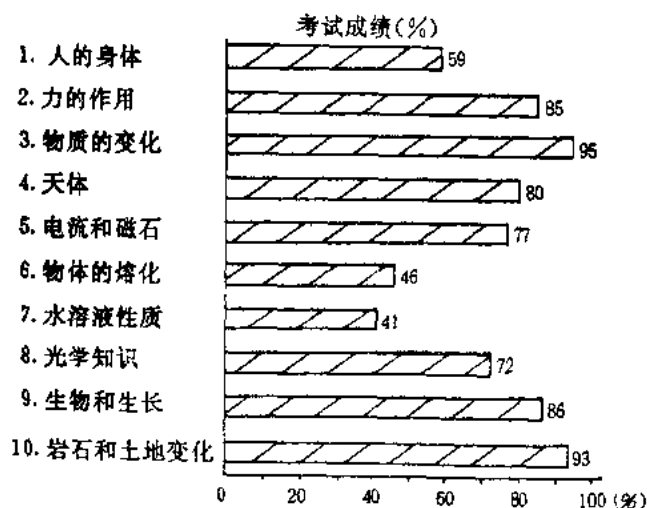


图1-1 模拟考试成绩示意图

图 1-2 把直方图和线图结合在一起。直方图表示的是相对频度。各单元按应花时间的多少从多到少顺次排列。线则表示累积的频度，从中可以看到随着单元数的增加累计应花的复习时间的增长趋势。很明显，“物体的熔化”、“水溶液的性质”和“电流与磁石”三部分应花费健二理科复习时间的75%左右。绘制这类图应注意，各单元所占相对频度之和应等于累积度数（均为100%），否则容易得出错误印象。

学习时间的分配

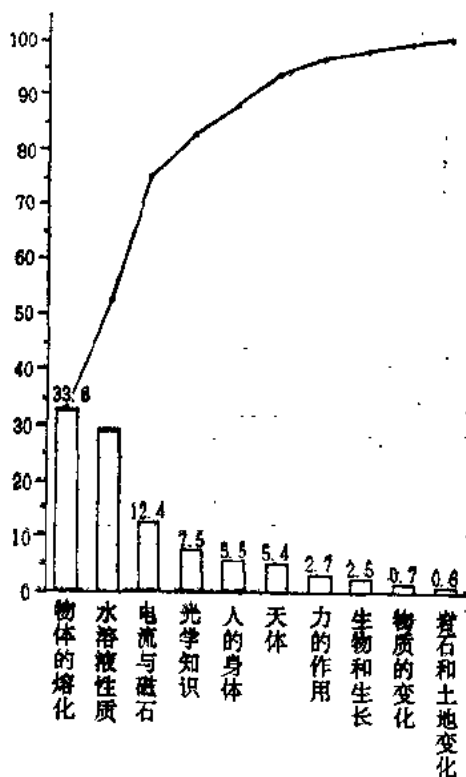


图1-2 学习时间分配相对频度和累积度数示意图

图 1-3 实即圆形构成比图，它以全圆的总面积作为 100%，各单元的构成比乘以 3.6 度即得到自身在圆内应占的圆心角度数。