

简 明 统 计 学

[英] H. T. 海斯莱特 著

暴奉贤 崔之庆 孙振声 等译

Q

A42

黑龙江人民出版社

1981年·哈尔滨

STATISTICS Made Simple
Advisory editor Patrick Murphy, M.Sc., F.I.M.A.
Made Simple Books
W.H.AILEN London
A division of Howard & Wyndham Ltd

简明统计学

〔英〕H·T·海斯莱特 著

暴奉贤 崔之庆 孙振声等译

黑龙江人民出版社出版

(哈尔滨市道里森林街42号)

黑龙江新华印刷厂印刷 黑龙江省新华书店发行

开本 850×1168 毫米 1/32·印张 11 8/16·字数 245,000

1981年5月第1版

1981年5月第1次印刷

印数 1—12,800

统一书号: 4093·42

定价: 1.30元

出版说明

本书是讲述数理统计理论和方法的普及读物，是英国伦敦艾伦公司发行的《简明丛书》中的一种。作者是英国 H·T·海斯莱特教授。这个译本，是根据艾伦公司 1976 年版本翻译的。

我们出版这本书，是为了适应我国四个现代化建设的需要，向广大经济、科技工作者以及经济管理和统计专业的师生，提供一本学习数理统计的参考书。

全书共十二章，系统而简明地介绍了数理统计的基本理论与方法，包括图形描述、位置量数、变异量数、基础概率与二项分布、正态分布、假设检验、相关与回归、置信限、非参数统计、方差分析等。内容浅显，说理明晰，文字通俗，举例较多，着重说明数理统计方法的应用，而对数学公式的证明则均从略。具有中学数学水平的读者都可以看懂。

这个译本，除将原书末尾名词索引的排列方法作了更动（原书索引按英文字母顺序排列，译文改为按汉字笔划排列）外，其他内容全部按原文原封不动地翻译过来。书中举例范围较广，例题较多，是其优点，但有些例题的选择，难免反映了资本主义社会生活的特点，希望读者在阅读和应用时加以注意。

参加本书翻译工作的有：暨南大学的暴奉贤同志和湖北财经学院的崔之庆、孙振声、颜日初、宁克庭、周兆麟、涂葆林、尚振礼等同志，由崔之庆、孙振声二同志复校，最后又经黑龙江省统计局史寅生、冯杞靖、李介民三同志校订。

前 言

在日常企业经营的各个方面，统计方法的应用日益显得重要。的确，想要举出一种不借助它本身所特有的统计数字来说明影响人类行为的活动，那几乎是不可能的。

但是，任何统计调查或市场分析之是否有用，完全取决于人们对他们所占有的资料加以说明的能力。由于现在没有一个商业企业愿意听取那些非专业人员根据粗略的估量所作出的模糊判断，所以对统计技术和方法有专门知识的人员的需要，比以往任何时候都更迫切，在对统计调查的结果担任着顾问职责的人的肩上，承担着很大的责任。这本书将使他们满怀信心地承担起这种责任。

在这本书中，海斯莱特教授假定它的读者除了中学代数以外，再没有其他数学知识。可能，总和的记号、不等式以及直线方程等对某些学生来说，是不很熟悉的，如果真是这样，他们将会发现，在他们需要说明的地方，问题都得到了论述，这就使本书内容十分完备。

书中有着大量的例题，但，由于统计的范围是如此广泛，以致不可能从每个不同的领域中都能举出例子。作者相信，使用本书的任何一个读者自己将会这样做。因为，他们在自己工作或学习的问题上是具有一定的知识的。读者在学习了本书所作的说明和例题以后，一定会以自己的具体运用来加以补充。后面几章问题比较复杂。对各种不同的统计检验将要逐步给以指导，并用一些例证来加以说明。很少有其他的书籍对这些检验作出这样详尽的研究。

第二、三、四章的内容是基本的，这些内容对学习全书都很需要。第五章的内容则是学习第六章所专门需要的，一般地说，概率的知识有助于理解全书(参见本丛书中的《简明数学》)，但它对使学生能完成各种检验来说，却不很重要。

本书包括对统计学基本原理的明晰表述，任何读者，不管他们从事什么专业，都会感到本书对他很有帮助。所以，对于想要取得中、高级水平的普通教育证书(G.O.E)而学习专业课程的学生，本书完全可以推荐给他作为学习参考书。

帕特里克 墨菲

目 录

第 一 章 什么是统计	1
统计的现时重要性	3
两种统计	4
第 二 章 资料的图形描述	7
引 言	7
随机抽样	8
资料的分组	9
频数分布和累积频数分布	13
资料的图形描述	15
直方图	15
频数多边形	18
累积曲线	18
习题	19
第 三 章 位置量数	22
引 言	22
全距中值	23
众 数	23
中位数	24
算术平均数	25
分组资料的中位数	27
总和的记法	31
分组资料的平均数	35
习 题	38
第 四 章 变异量数	40
引 言	40

全 距.....	42
平均绝对差.....	43
方差和标准差.....	45
分组资料的方差和标准差.....	52
习 题.....	58
第 五 章 基础概率和二项分布.....	59
引 言.....	59
简单事件的概率.....	61
两个事件的概率.....	65
三个或更多事件联合的概率.....	75
排 列.....	82
组 合.....	87
其它概率.....	92
二项分布.....	93
二项分布的理论平均数	107
二项分布的理论方差	112
习 题	116
第 六 章 正态分布	119
引 言	119
正态分布	121
标准正态表的用法	123
非标准正态概率	131
正态近似二项式	135
习 题	145
第 七 章 一般的统计假设检验	147
引 言	147
统计假设以及两类错误的性质	148
对 $H_0: \pi = \pi_0$ 和相对的特定的择一假设的检验	152
关于正态分布平均数的检验	161

习 题	165
第 八 章 进一步的统计假设检验	167
引 言	167
检验 $H_0: \mu = \mu_0$, 正态总体, σ^2 未知	168
关于非正态总体平均数的检验	173
关于两个比率差异的检验	177
关于两个平均数差异的检验	185
习 题	194
第 九 章 相关与回归	196
样本相关系数	196
r 的计算法	199
关于总体相关系数假设的检验	201
线性回归	207
用最小二乘法求回归线	208
在回归问题中检验关于 μ 的假设	214
在回归问题中检验关于 β 的假设	218
习 题	221
第 十 章 置信限	223
引 言	223
μ 的置信区间	226
π 的置信区间	233
$\mu_1 - \mu_2$ 的置信区间	236
$\pi_1 - \pi_2$ 的置信区间	243
ρ 的置信区间	246
习 题	248
第十一章 非参数统计	250
引 言	250
χ^2 分布	250
列联表	261

等级相关系数	266
符号检验 (一个总体)	272
威尔柯克斯符号—秩检验	276
秩和检验 (两个总体)	281
习 题	287
第十二章 方差分析	290
引 言	290
一向方差分析	291
一向方差分析——另一种方法	300
样本容量不等的一向方差分析	307
二向方差分析	315
习 题	321
附 录:	
主要符号一览表	323
计算用表	
平方表	324
平方根表	326
t 分布表	330
标准正态分布面积表	331
χ^2 分布表	333
菲希尔—Z 数值 (Z_r) 表	334
斯庇尔曼等级相关系数 (r_s) 表	335
威尔柯克斯符号—秩数值表	335
秩和临界值表	336
F 分布表 $\alpha = 0.05$	337
F 分布表 $\alpha = 0.01$	339
F 分布表 $\alpha = 0.025$	341
F 分布表 $\alpha = 0.005$	343
习题解答	345
索 引	353

第 一 章

什 么 是 统 计

为了通俗易解地学习统计这门学科，我们必须首先弄清“统计”一词在今天的意义和关于它的起源的一些情况。

如同许多其它的词一样，不同的人对“统计”一词有不同的理解。有些人一听到这个词就想到关于出生、死亡、结婚、离婚、车祸等数字表格。这的确是对这个词的生动而又适当的使用。事实上，“统计”一词首先被用于国家事务，用于政府为有效地实施计划、管理和征税的需要而收集资料的任务。收集和分析这些资料的人曾经被叫作“统计员”(Statists)。这比原来以了解国情为职业的“统计学家”(Statistician)一词更加明确。

当然，今天“统计”一词在第一种意义上几乎是指任何用数字所提供的真实资料——即所谓“事实和数字”。电台和电视台的广播员告诉我们他们“在几分钟内将有统计节目”。报纸上经常发表有关美女竞选和提供竞选统计的文章。

可是，“统计”一词还有另外的意义。没有研究过这门学科的人对它是比较生疏的。统计是应用数学领域中的一门学科，有它自己的符号、术语、内容定理和技术方法。当人们学习统计学这门学科时，他们是想要精通这些技术的某些方面的。

对那些已经了解到统计科学奥秘的人来说，“统计”一词还有第二种意义。在这种意义上，“统计”(Statistics)是指从样本资

料中计算出来的数量群。这样计算出来的单一数量叫做“统计量”(Statistic)。例如样本平均数是统计量,样本中位数及样本众数也同样是统计量(它们将在第三章中论述),样本方差和样本全距是统计量(二者在第四章中论述),样本相关系数(在第九章中论述)也是统计量,等等。

我们总括一下“统计(Statistics)”一词的意义是:

1. 众所周知的意义是事实,数字,图形和表格。在这个意义上,这个词是复数。

2. 学科本身的意义是这门学科有它自己的术语、方法论和它自己的知识总体。在这个意义上,这个词用作单数。

3. 从样本资料推算出来的数量群。在这个意义上,这个词也是复数。

在本书中,我们将不用上述“统计”一词的第一种意义,当我们要说到事实和数字时,我们将用“观察值”或“资料”这个词。我们将偶然把从样本资料中计算出来的数量称为统计量(Statistic)。在这些情况下,我们将用上述第三种意义上的“统计”一词的单数形式。当我们用“统计”(Statistics)一词时,几乎总是指这门学科本身,指它的知识总体。

统计方法论有时被误解了,以致产生了许多关于统计和统计学家的怪论。例如“统计学家是一个从无根据的假设到预料中的必然结局之间划一条精确的界限的人”。这就产生了对统计方法的滥用,虽然不是被职业的统计学家所滥用。应用大多数统计方法,一定的假设是必需的,假设的数量和范围则根据情况而异。也许有些人提出他们明知是不能证明的假设或伪装怀疑。也许有些人事先已经作出了结论,却再去挑选他们的样本或篡改数据来证明他们的结论。故意这样滥用的人是不老实的。

反对统计方法和方法论的人说:“统计分析经常意味着对有

分歧的数字进行篡改，用意义含糊不清的办法去解决不明确的问题”。

可能为大家所熟知的评论，是马克·吐温对迪斯雷利说过的一句话：“有三种谎言：谎言、糟透了的谎言、统计。”也有另外的众所周知的评论，用以批评有人应用统计的态度：“他使用统计就如同一个醉汉使用街灯一样，是用来支撑他，而不是用来照明。”

统计的现时重要性

统计方法的应用是这样普遍，在我们的生活和习惯中，统计的影响是这样巨大，以致统计的重要性无论怎样强调也不过分。

我们现代农业之所以能够丰产，部分地应归功于把统计应用到农业实验的设计和分析上。这是统计方法应用得比较早的一个领域。有些问题要靠统计方法去帮助解答：哪一种玉米收成最好？哪种混合饲料喂养小鸡将获得最高体重？哪一种混合草种每亩能收获更多的干草？所有这些问题和其它许多问题，通过本地超级市场，我们所有的人都受到直接的影响。

统计方法也经常用于医学和药学的研究上。通过首先对动物然后对人类进行实验的方法决定新药的效果。医学的新发展和新药物对我们大家都有影响。

统计也同样被政府应用。经济资料被用来研究和影响政府在税收、公共工程（例如道路、桥梁等）的基金，公共救济基金等领域方面的政策。失业统计影响着降低失业率的努力。统计方法被用来评价各种军事装备（从手枪、子弹到庞大的导弹）的性能。概率论和统计（尤其在一个称为统计决策论的更新领域）被用于

帮助人们在最高水平上作出非常重要的决定。

统计在私人工业中的应用几乎和在政府工作中的应用同样重要，作用同样广泛。统计方法被用来对正在生产的产品进行质量控制，并且对新产品在其销售以前作出评价。统计还被用于销售，用于决定扩大营业，用于分析广告效果等等方面。保险公司应用统计资料在实际水平上决定保险率。

还能继续列举下去。统计被用在地质学、生物学、心理学、社会学，用在那些必须在不完整的资料的基础上作出决定的任何领域。统计被用在教育试验和保险工程方面。在气象学、在天气预报科学中，现在统计正在被广泛应用。

另外，统计还研究在月圆时捕捞鲜鱼的效果。还研究对两种玻璃杯，餐馆用哪一种比较好。还研究在玩一场凭技巧和碰机会的游戏中（如玩桥牌时）最适宜的策略等等。

对统计和统计方法的效果，我们每个人都很少怀疑。在我们的工资袋中，在我们的国防安全上，在我们的保险费中，在我们对各种产品的满足上，在我们的健康等方面，统计研究的结果是看得见的，但很可能没有被认识到。

两 种 统 计

除简要地研究概率的基本原理外，在本书中有两种统计的论述。在第二章、第三章和第四章中，我们主要关心的是资料的描述。在第二章中，我们论述了资料的图形描述。在第三、四章中，我们论述了资料的数字描述。这种统计的当然名称是**描述统计**。资料的分组，资料分组后的结果绘制符合频率分布的直方图，用别种图形对数字资料的描述，如直线图、条形图和图画，样本

平均数，中位数或众数的计算，方差、平均绝对离差和全距的计算。描述统计就是论述这些问题。回溯到统计的产生，在 19 世纪和本世纪的初期，主要的是描述统计。

第二种重要的统计叫做**推断统计**。把统计说成是面对不肯定的事物作出决定的科学，那就是在不完全资料的基础上作出比较好的决定的科学。为了作出一个有关总体的决定，样本（一般抽取少数单位）是从那个总体中挑选出来的。抽取样本通常是随机进行的。虽然有多种多样的抽取样本的方法，我们在本书中采用的方法叫做随机抽样法。正如这个名称所表明的，随机抽样是这样一种抽样，它的样本单位是用某种不受试验者控制的方法进行抽取的。对随机抽样有各种各样的数学定义。我们考虑的是作为一个样本每个总体单位都有同等的机会被抽选出来，而且选任何一个单位作为样本都不影响选其它单位作为样本。

我们在随机抽样的基础上推论有关总体的情况。这种在抽样基础上对总体的推论就叫做**统计推断**。换句话说，统计推论是用样本对从中抽出样本的那个总体做出结论。

让我们举出统计推论的几个实例。假如一个三轮车制造者买了大量的螺栓，如果有缺陷的螺栓超过了 3%，三轮车制造者可以有权拒绝接收这批货物。当然在使用以前对全部螺栓都进行检查是不行的，那花费的时间太多了。也不能在装配三轮车的时候，碰到有缺陷的螺栓就把它放在一边。因为这些螺栓在使用以后就不能再退回去了，即使有缺陷的螺栓达到 20% 也不能再退了。三轮车制造者当然不肯用一批废次品占很大比重的螺栓，因为已经知道它有缺陷，然后再做一批令人满意的螺栓来换掉它是划不来的。基于以上几个理由，三轮车制造者必须有一个快速的、廉价的方法，用以了解这批货物是否废次品太多。所以他从这批交付的螺栓中随机抽取一个样本，从样本中废次品所占的百分比

推算出在总体中（整批货物）废次品所占的百分比。这是一个统计推论的例子。

再举一个统计推论的例子。一个医学研究工作者要确定一种新药是不是比老药优良，在医院里把一百个病人随机划分为两组，一组用老药，另一组用新药。从开始用药的那天起，从每个病人那里，取得各种医学资料。测验十天以后，对这两组资料进行分析和比较，就可得到有关这两种药的效力的结果。

类似的例子将从第七章开始详细讨论。在那里将首先讨论统计假设检验问题。

统计假设检验的程序通常是：提出一个关于总体的假设，一般称为零假设。从这个总体中抽取一个随机样本，再从样本资料中计算出通称为统计量 (Statistic) 的数量。这个零假设是接受或是抛弃，取决于统计量 (Statistic) 的数值（“择一假设”是与“零假设”同时形成的，拒绝“零假设”就意味着自动接受“择一假设”。）这样，统计假设检验就是统计推论的一个例子。因为用样本作出了有关总体的决定。第七、八、九、十一章和第十二章是有关（部分或全部）统计假设检验因而也是统计推论的问题。第十章讨论了一个重要的问题，叫做置信区间，也涉及到统计推论。

总之，在本书中，这门学科自然而然地归纳为三个方面。第二、三章和第四章讨论资料的描述，或用图表或用数字表示，它们被归纳为描述统计。在第五章和第六章中讨论概率论的某些很简单的问题：基础概率、两个重要分布——二项分布和正态分布——两者的联系以及正态表的使用。最后六章主要是有关统计假设检验的问题，这一部分的学科内容称为推论统计。由开头六章组成的前面两类的内容是后一类内容的准备。

第 二 章

资料的图形描述

引 言

本章介绍样本资料的表述问题。在论述资料的分组和绘制直方图以前，我们将简要地讨论一下随机样本的概念以及如何抽取随机样本的问题。

如果一个样本是从无限个单位组成的总体中抽取出来的，抽取的方法是：从总体中抽取任何单位都不影响总体中其它单位的抽取，并且每个总体单位都有包含在样本中的同等的可能性，这就叫做**随机样本**。如果一个样本是从有限总体中抽取出来的，抽出后再放回去（某个总体单位被抽出以后，又归还到总体中，可以不止一次被抽出来）。这就是随机样本的确切解释。换句话说，抽出以后又放回去的有限总体，在理论上可以视同一个无限总体。

如果一个样本是从有限总体中抽取出来的，抽出以后不再放回去（这个单位抽出观察以后不再放回到总体中去）。如果同样大小的其它样本都有同等的被抽选的机会，那么我们仍然说它是一个随机样本。没有一个样本比其它样本有更多的被抽选的机会。例如，从一个装有一百张彩票的容器中一次抽出三张彩票，就是一个从有限总体中无退还抽取的例子。它是随机样本，因为任何

其它三张彩票都有同样的机会被抽出来。

“随机”一词意味着样本是按照这样一种方法进行抽选，以致于它不可能预言那个总体单位将包括在样本内，任何一个单位被抽出来仅仅是一个碰机会的问题。为了应用本书中所论述的统计技术分析样本资料，样本必须是随机的（很少例外）。统计方法已被统计理论所证明，而统计理论又建筑在概率论的基础上，而在概率论可以应用之前，我们必须有随机样本。

随 机 抽 样

抽取一个随机样本有时不是一件容易的事情。如果总体很小，抽取样本的最简单的方法是把总体单位列出（例如写在几张小纸片上），然后“从一顶帽子中”进行抽选。

当能用整数来计量每个总体单位时，就可以用**随机数表**来抽取随机样本。随机数表是将阿拉伯数字列成表，通过随机过程来抽取。给各个总体单位分配一个整数的方法是简单地将各单位列成 1、2、3 等等。（当总体单位不便列数的时候，就可用另外的方法，通过随机数抽取随机样本。）每个总体单位在随机数表上有一个相对应的数字（或许多于一个相对应的数字）。要抽取一个随机样本，我们就要在随机挑选的地方，从随机数表上开始读数，对于每个所读到的数字，其相应的总体单位就包含在这个样本中。例如，如果我们的总体由一千个单位组成，它们的数字分配是从 000 到 999。如果我们在随机数表上读到数字 027，831 和 415，我们就将号码为 027，831 和 415 的总体单位包括在随机样本之中。

表 2·1 是一百个学生在一次智能测验中口试部分获得的分数