

统计学入门

〔英〕亚瑟·哈蒙德·霍尔 著

暴奉贤 陈维中 郭 祝 谭达才 译

龚 鎰 尧 校阅

立信会计图书用品社

统计学入门

[英]亚瑟·哈蒙德·霍尔 著

暴奉贤 陈维中 译
郭 祝 谭达才
龚 鑑 尧 校阅

立信会计图书用品社

封面设计：范一辛

统 计 学 入 门

〔英〕亚瑟·哈蒙德·霍尔 著

暴奉贤 陈维中 郭 祝 谭达才 译

立信会计图书用品社出版发行

（上海中山西路 2230 号）

新华书店经销

立信梅李印刷联营厂印刷

开本 850×1168 毫米 1/32 印张 5.75 插页 2 字数 140,000

1983 年 12 月第一版

1988 年 10 月新一版 1988 年 10 月第一次印刷

ISBN7-5429-0041-2/F·0041

定价：1.89 元

译 者 的 话

本书为英国统计学家亚瑟·哈蒙德·霍尔所著，伦敦麦克米兰公司一九七八年出版。全书分两个部分，共十九章。其内容主要阐述数理统计的一些基本原理及其在社会经济活动中的应用。本书通俗易懂，例证丰富，每章都附有习题，并在书后附有习题答案。具有中学文化水平即可自学。它对于我国综合与专业系统各级统计人员以及财经院校师生了解西方国家统计理论和经验，也有一定参考价值。

本书由下列人员翻译：

暴奉贤——前言，第十二、十三、十四章，习题答案；

陈维中——第一、二、三、四、五章；

郭 祝——第六、七、八、九、十、十一章；

谭达才——第十五、十六、十七、十八、十九章。

全书译文由龚鑑尧校阅。

一九八三年六月

DAZ 22/06

前 言

本书是一本在莱撒姆圣安尼进修学院用作普通国家毕业考试和普通级毕业考试的补充教材。商业和营业学系主任达灵顿先生经常给作者的帮助和鼓舞,谨在此表示衷心感谢。

近年来随着对统计学兴趣的高涨,经常出现一些本来对其他主题感兴趣或是注重搞理论的人,也教起统计课程来了。希望在本书的例子和作业中广泛利用的统计资料会对他们是有价值的。普通国家毕业考试的技术顾问们已及时指出了在给他们的试卷中的适当资料的有限用途。而主考人一般都试图扩大他们试卷的范围,以尽力鞭策单纯死记硬背的人!

在本书中将不会有实际的试卷,但是有许多比较难的问题。由于委员会以及普通国家毕业考试之间的分歧,习题已经大大地增多了。补充课程内容和对以前的试题进行适当的修订,这对于教师和学生双方都是有益的。

数学和统计用表没有包括进来。只有前后几页不便去掉,因为学生们自己要熟悉这些表,他们在考试时要用到。当然,用袖珍计算器来代替表的日子是会到来的。

教师和学生在他们自己的教学过程中,将发现这本教材十分灵活,并可以调整论题的次序以适合他们自己的需要。两个部分中的每部分都是一个逻辑系统,但不是完全独立的。不过,第十八章至少需要在第十四章之前学习。

作者感谢中央统计局局长和皇家文书局的管理员,由于他们允许应用《统计摘要月刊》(M.D.),《年度统计要览》(A.A.),《地区统计要览》(A.R.)和《支付决算蓝皮书》(B.B.)中的政府统

计资料。对于他们允许在最后一章中提供政府统计方法选编摘要,也在此一并致谢。

亚瑟·哈蒙德·霍尔

伦敦·英国

目 录

第 一 部 分

第 一 章	绪论——怎样学和为什么要学统计学	1
第 二 章	平均数	4
第 三 章	算术平均数	7
第 四 章	离势	16
第 五 章	累积次数	27
第 六 章	间断量和连续量	33
第 七 章	众数和中位数	42
第 八 章	近似值	51
第 九 章	双变量分布——相关和回归	59
第 十 章	等级相关	71
第十一章	时间数列	77
第十二章	加权平均数和指数	90
第十三章	统计资料的各种图示法	105
第十四章	正态分布	119

第 二 部 分

第十五章	抽样	128
第十六章	调查和民意测验	134
第十七章	资料的制表	141
第十八章	统计图	149
第十九章	政府统计和其他统计	163
附录: 习题答案		171

第一部分

第一章 绪论——怎样学和为什么要学统计学

这本书分为两个部分：第一部分论述必需的数学和计算，因而可以说是数理统计；第二部分论述背景和说明一些需要统计学的情况，这部分通常称为描述统计学。

希望每个学生学完这个课程后，不仅会成为一个统计数字的操作者，而且会成为一个统计数字的忠实而熟练的解释者。

但是，我们也常常会看到，有些人缺乏清楚的理解能力。许多广告都是基于普通人们的轻信，使他们很容易接受这样一些说法：

六个人中的五个人都说不出这个差别……

大多数家庭主妇喜欢……

去年工资(价格)已提高……

平均容量……

有时候，某人进行一次调查，所得到的结论是以上述那样一些说法为依据的。我们不知道的是：多少个六个人被询问过；大多数指的正好是全部家庭主妇的 51%，或者是某一些家庭主妇的 99%；或是谁的工资提高了？以及什么是平均容量？

我们已经面临着一些统计人员的问题。如果我们不能询问全部家庭主妇，那么，我们应询问多少并怎样决定应询问那些家庭主妇的工作。或是，某个制造商怎样才能断定一盒火柴的平均容量是 49 根？这就把我们引导到以下各页论述的许多统计概念中的两个概念——样本的概念和平均数的概念。家庭主妇和火柴盒两者的数量是那么多，以致不管哪一种都不可能全部进行调查，而需

要从总体中抽选一些来进行调查。所有包含在调查中的这些事物或人构成一个“总体”，从他们中抽选出来的一些事物或人构成一个“样本”。平均数的概念是一个非常熟悉的概念，但它象总体和样本一样，涉及的问题很多，本书对每一个这样的问题都必须在一定的深度上进行探讨。

调查不是要得出受到公众注意的说法，以便去影响街上男人

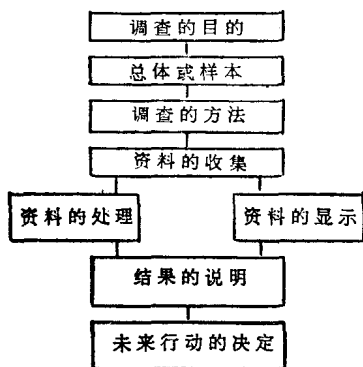


图 1

或女人的决心。在工业方面有许多说法是会影响到的产品的质量、生产效率和未来的计划等等的。统计人员的整个工作程序见图 1。

最后一项可以转回到头一项，然后再进行。初露头角的统计人员的想法是美妙的。如果他是一个好的统计人员，他决不会找不到工作！

习 题 1.1

1. 列举一些由电视广告员广播的稿子，并找出每篇中一些值得批评的地方。

2. 人口普查有些什么目的？

3. 列举一些经过检验就会被破坏而无用的产品。在这样的情况下，应对总体进行检验还是抽选样本来检验？

4. 决定正在计划建设的一个新的大超级市场里需要多少个结帐点，要进行什么样的调查？

5. 就你力所能及去调查许多家庭购买的东西，并将它们分为：

① 标明“平均含量”的。

② 没有标明含量的。

③ 正好说明含量的。

6. 列举一些使用平均数的娱乐活动，并找出它们使用平均数的理由。

7. 设计一张可以由下列各种人填写的调查表：

- ① 医生；
- ② 农民；
- ③ 教师；
- ④ 政治家；
- ⑤ 油漆制造商。

8. “雪白”；“空前大减价的电子计算机”。一个统计员是不可能作这样的陈述的。为什么？尽可能从你家里的报章杂志或电视中找出许多类似的令人讨厌的广告内容，并给予批评。

第二章 平 均 数

一般使用“平均数”一词，要比通常在学校里讲授的意思广泛得多，而统计学家为了不致影响他的结论，必须有十分精确的定义。

一群儿童某次测验获得如下的分数：

1、2、3、3、3、3、4、5、6、7、9、9、10

对于这组数，我们能够计算三种平均数值。它们是：

1. 算术平均数，是把这十三个数字加总然后除以 13。这个数值是 5。

2. 中位数，是把这些数字按大小顺序排列居于中间位置的数。这里的中位数的数值是 4，有六个分数小于它，六个分数大于它。

3. 众数，它是出现次数最多的数值。这里的众数是 3，出现了四次。

算术平均数、中位数和众数称为“集中趋势值”，因为其它数值都趋向于群集在每一个这样的数值周围（但对于众数来说，有可能是最大的或最小的数值）。亦有称之为“位值”的。

算术平均数的计算需要把全部的数字加总，而不仅仅是将不相同的数字——1、2、3、4、5、6、7、8、9、10 相加。算术平均数的一个十分重要的性质是答案能够验证。如果算术平均数（简称“平均数”，这是统计人员最常用到的一种平均数）分别与每一个数值相减，我们就得到所谓“对平均数的离差”。在上面的例子里对平均数的离差是：

-4、-3、-2、-2、-2、-2、1、0、1、2、4、4、5

得到的总和是 $-16 + 16 = 0$ 。

对算术平均数的离差之和永远为零!

中位数在数字不多并且已经按大小顺序排列好时, 很容易找到; 不然的话, 就很麻烦, 并且不容易验证。

众数需要计算每个不同数值出现的次数, 并且存在着与上述中位数相同的缺点。

这三种平均数都将在后面各章中作更充分的论述。

习 题 2.1

1. 计算 2、4、6、8、10 这些数字的算术平均数。写出它们对平均数的离差并验证离差的总和是零。

2. 2、4、6、6、8、8、8、10、10 这些数字的算术平均数是什么? 并用它们对平均数的离差进行验证。

3. 第 1 题中各数的中位数是什么?

4. 第 2 题中各数的众数是什么?

5. 2、4、6、6、6、8、10 的算术平均数是什么?

6. 第 5 题中各数的众数是什么?

7. 第 5 题中各数的中位数是什么?

8. 一组数对数值 8 的离差是 2、3、-1、0、4、-2、-2、-2、-2, 其算术平均数是什么? 计算出这组数, 并求出它们的中位数和众数。

9. 一组数对平均数 17 的离差是 3、-4、5、1、-2、0 和 x 。 x 的数值是什么?

10. 下面是 9 份一磅重的马铃薯, 每份的马铃薯数分别为: 3、9、4、7、5、3、7、3、4。求出它们的算术平均数、中位数和众数。

11. 十个男孩法语成绩的算术平均数是 8 分, 他们的总分数是多少?

12. 在第 11 题所说的男孩的同一个班里, 20 个女孩法语的平均分数是 5 分。整个班 10 个男孩和 20 个女孩的平均分数是多少?

13. 某班男孩和女孩共 25 人, 算术课平均成绩是 12 分。该班 15 个男孩的平均分数是 14 分, 女孩的平均分数是多少?

14. 五个学生的代数课分数是: x 、 $3x$ 、 $2x-1$ 、 $5x+9$ 、 $4x+7$ 。其平均分数是多少?

15. 某组有 n 项, 算术平均数是 m ; 另一组有 N 项, 算术平均数是 M 。证明其合并组的算术平均数是

$$\frac{nm + NM}{n + N}。$$

16. 在某老人院里 11 位百岁以上老人的岁数是 101、102、103、104、108、102、105、102、110、105、102。下面说法中哪些是正确的？

- ① 算术平均数 = 中位数 = 众数；
- ② 众数 > 中位数 > 算术平均数；
- ③ 中位数 > 算术平均数 > 众数；
- ④ 算术平均数 > 中位数 > 众数。

17. 在某个技术工厂特殊元件的 10 个样本里，废品的数字是 4、3、7、5、4、3、6、2、3、4。如果每个样本包含 100 项，合格品数的中位数是什么？

18. 在回答测验时，答错的扣 1 分，答对的给 1 分，不回答的为零分。有 5 个答错，15 个答对，5 个没回答，其平均分数是多少？

19. 有 6 个数，对数值 10 的离差分别为：-4、+2、-3、+1、0、+10。这 6 个数字各是多少？它们的平均数值是什么？

20. 有 10 个数、对数值 6 的离差分别为：-3、-2、-2、-1、0、0、4、4、5、5。下面说法中哪些是正确的？

- ① 其中有些数是负数。
- ② 全部数都是正数。
- ③ 这些数的平均数是零。
- ④ 这些数的平均数是 1。
- ⑤ 这些数的平均数是 7。

第三章 算术平均数

§ 3.1 在统计人员看来，计算是很平常的事，通常要有一部台式计算机或者袖珍计算器。而许多问题比人们想象的还要容易些，本章将介绍这种方法。

4、5、7 和 8 的算术平均数是 6。

104、105、107 和 108 的算术平均数是 106。

这好象完全是一种逻辑上的升级。由于在第二组数字里每个数字都增加了 100，因而自然地设想到这个平均数也会增加 100——而这种推论是正确的。

52、52、53、54、59 的平均数是什么？答案是 54、它比 2、2、3、4、9 的平均数多 50。

648、597、589、604 的算术平均数是什么？这个答案一定是接近 600。因此，如果我们对这四个数中的每一个数都减去 600，就是 +48、-3、-11、+4、求出它们的算术平均数，然后再加上 600，这样就得到答案：

$$9.5 + 600 = 609.5$$

我们减去的这个数字有好几个名称——虚拟平均数、假定平均数、虚拟原点、假定原点。注意，这些新的数字——叫做离差，因为它们原来是数字与假定平均数之差——是正数或者负数都没有关系。如果它们的总和是正数，我们所计算的平均数就会大于假定平均数；如果它们的总和是负数，我们所计算的平均数就会小于假定平均数。用 31、77、23、21，这组数字和假定平均数 40 来验证上面这段陈述，则这些数对 40 的离差是 -9、37、-17、和 -19，而它们的平均数是 -2。我们所要计算的算术平均数是

$$-2 + 40 = 38。$$

习 题 3.1

1. 用：①常规方法；②取对假定平均数 60 的离差，计算 54、61、93、104、75 和 99 的算术平均数。

2. 如果 654、479、991、863、275 和 338 这些数字的平均数是 600；454、279、791、663、75 和 138 的平均数是什么？什么样一组数字会得到平均数 725？

3. 取自某车间的样本的九个原件按毫米计测定的长度是 4.9、5.4、5.2、4.7、4.6、5.1、4.8 和 4.2。取假定平均数为 5.0 毫米，并计算它们的算术平均数。

4. 求 13、21、19、22、34、15 和 16 的算术平均数。写出下面各组数字的平均数：

① 3、11、9、12、24、5 和 6。

② 15、23、21、24、36、17 和 18。

③ 1.3、2.1、1.9、2.2、3.4、1.5、和 1.6。

④ 26、42、38、44、68、30 和 32。

5. 称量七个样本按克计的结果是：6.45、6.51、6.43、6.49、6.50、6.40、6.47。取假定平均数为 6.40，求它们的算术平均数。

§ 3.2 从上个习题的问题中，将引伸出一些新的原理。它们是：

(1) 一组数的每个数字都增加(减少)同一的数量，其算术平均数等于这组数的算术平均数增加(减少)同一的数量。

(2) 一组数的每个数都乘(除)以同一的数量，其算术平均数等于这组数的算术平均数乘(除)以同一的数量。这在采用不同的计量单位——毫米代替分米，百分之一英寸代替英寸，等等较为方便时，是非常有用的。以习题 3.1 的第 5 题为例，那组数对 6.40 克的离差用百分之一克作为计量单位就是 5、11、3、3、9、10、0 和 7。

综合(1)和(2)我们能够系统地阐述为：

(3) 如果一组数用四种算术方法中的任何一种(加减乘除的数字为常数)改变为另一组数，则用同一的常数、同一的顺序、同一

的算法对原来一组数的平均数进行计算,就会得到这个新的一组数的平均数。

用下面的数值验证:

(1) 2、3、7。

对每个数值都乘以 3 然后加 5,就得到新的数值 11、14、26。

原来的平均数是 4,而新的平均数是 $17 = 4 \times 3 + 5$ 。

(2) 19、29、54。

对每个数值都减去 4 然后除以 5,就得到新的数值 3、5、10。

原来的平均数是 34,而新的平均数是: $6 = (34 - 4) \div 5$ 。

§ 3.3 对于一个大的分布(一个分布是一组数值、度量或者数字),其平均数的计算是较为麻烦的。留心前面所述的论点,注意精心和准确,这是很普通的事情。有许多数字,特别是那些超过一定范围的数字,必然会有一些数值多次重复出现。这许多数字需要制成次数分布表,次数(f)是指某个数值出现的次数。

对十打火柴进行检验,看看制造商声称“每盒火柴平均容量为 49 根”是否有根据。次数分布是每盒火柴根数在 45 至 52 之间:

容量	45	46	47	48	49	50	51	52
盒数	5	6	13	28	37	22	4	5

求平均数,每对数字相乘并加总,得到火柴根数总和,然后除以 120 就得到答案。在采用假定平均数时,这个规定同样要遵守。取假定平均数为 49(因为 49 是制造商声称的数值)并采用竖式安排,因为这样做对我们更为方便些。

看来制造商声称的是有根据的——49 是最接近实际的一个数字。

在这里可能会提出一个问题,“为什么要这么麻烦地算出一位小数?不管怎样你都不可能有一根火柴的一部分。”小数的用处是使我们能够说这个平均数并非恰好是 49;在 10 盒火柴里你可以预期得到 486 根火柴,而不是 490 根,要少四根。从另方面看,这

每盒火柴根数 x	盒数 f	对 49 的离差 d	fd
45	5	-4	-20
46	6	-3	-18
47	13	-2	-26
48	23	-1	-23
49	37	0	0
50	22	1	22
51	4	2	8
52	5	3	15
合计	120	-4	45-92=-47

$$\text{平均离差} = -\frac{47}{120} = -0.4 \quad \text{平均数} = 49 - 0.4 = 48.6$$

个表显示出按制造商声称的平均数计算, 这 120 盒火柴就少了 47 根, 相当于另一盒火柴的根数。如果制造商在 120 盒火柴里就有一盒是空的, 那他整年销售所得的利益是多少?

注: 用 x 表示一组数的平均数通常写为 $\bar{x}$, 读作“ x 靶”。

用 x 表示一组数的总和通常写为 Σx , 读作“西格马 x ”在这个例子里

$$\Sigma f = 120, \quad \Sigma fd = -47.$$

习 题 3.2

1. 在某月里每周从 26 个中心站买莴苣, 价格的分布如下所示。

价格(便士)	5	6	7	8	9
莴苣数	35	48	10	8	3

计算每个莴苣价格的平均数。

2. 150 瓦灯泡单价的分布如下:

价格(便士)	11	12	13	14	15	16	17	18	21
灯泡数	5	22	13	9	21	20	22	17	1

计算每个 150 瓦灯泡的算术平均价格。

3. 在第一学期初和第一学期末对 x 学校一年级某学生小组进行算术测