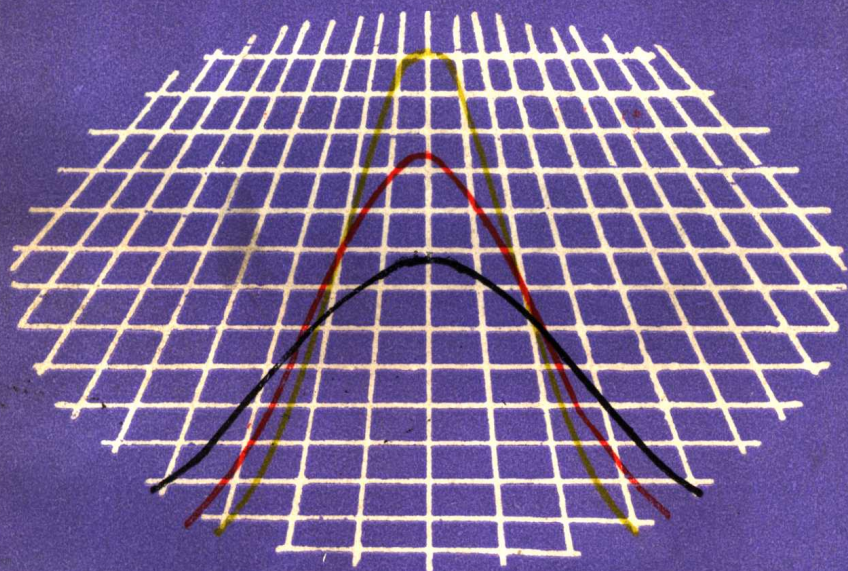


管 理 统 计

杨喜寿 著



山 东 大 学 出 版 社

管 理 统 计

杨喜寿 著

山东大学出版社

管 理 统 计

杨喜寿 著

※

山东大学出版社出版

山东省新华书店发行

山东大学印刷厂印刷

※

850×1168毫米 大32 12印张 312千字

1991年6月第1版 1991年6月第1次印刷

印数 1—2000册

ISBN7-5607-0455-7/O·32

定价3.70元

序

近数十年来，统计学已深深地扎根于经济与管理中，并且硕果累累。

当今有一种流行的看法：“世界正进入以信息为中心的时代。”许多情况下，信息是包含在大量的数据之中的。有效地收集、整理数据，从数据中提取有用的信息，从而推断、预测某些社会、经济现象（包括宏观的和微观的）的发展变化规律，以至于作出科学的决策，这些就是管理统计的主要内容。

关于数据的收集，包括两个方面的内容。其一，根据不同的目的，设计出表达各种现象的指标，以至建立一整套合理的、严密的指标体系。这方面的内容包含在相应的各种社会经济统计学中，如国民生产统计、劳动工资统计、商业物价统计、财政金融统计等。本书不讨论这方面的内容。其二，设计满足推断精度要求、省时、省费用的收集数据的方法。本书第九章抽样调查方法主要研究这方面的问题。

收集到的原始数据，常常是杂乱无章的。为从其中提取有用的信息，需加以整理，例如列表、画图，计算平均数、比例数、趋势函数等。这些又称为描述统计。本书前三章主要介绍这方面的内容。

统计学更为重要的任务是利用对一个总体来说不完全的信息来对该总体的某些特征作出判断。例如，有10 000发炮弹（总体），我们要判断其中的废品率（不能爆炸的比数）。从10 000发中任取20发作发射试验，若其中1发不爆炸。这是不完全的信息。完全的信息，需将全部10 000发炮弹都试验。如何利用这种不完全信息对总体特征（例如废品率是多少？是否可信不超过

5%?等等)作出判断,这类问题构成了推断统计学。本书第六章到第十章都属于推断统计学内容。

统计学作为方法论科学,一个突出的特点是应用上的广泛性。它在社会科学、自然科学、工程技术等几乎各个领域都有用武之地。

本书在介绍统计方法时,强调其应用性及应用的广泛性。许多重要的概念、方法、原理都是结合实例来讲解。这些实例来自管理的许多不同的领域。其中有一些是作者自己研究工作的一部分。例如,例3.3.1是关于全国电风扇销售量的预测;§6.7(三)和例8.3.2是教育管理方面的;例8.6.2研究了农业劳动力投入的效益问题;例8.7.1是关于消防管理工作的;例9.6.5是一个社会敏感问题的调查分析;等等。

统计学虽说应用极其广泛,但统计方法本身其骨架相当严密、清晰。

本书第四、五两章是概率论的基本知识。是以后几章所介绍的统计推断方法的理论基础和工具。学习这两章需要一元、二元函数微积分的知识。这两章所讲的概率论,比一般数学类概率统计书中所讲的内容要浅一些,主要表现在没有详细地介绍极限定理。但较之一般管理统计类书又深了一些。为了不至于使读者陷于数学困境,一些较难的推导过程都写在附录里。这是为那些对理论感兴趣而又希望深入研究的读者准备的。

第六章主要介绍关于总体特征数的点估计、区间估计、假设检验等最基本的统计推断方法。这些既是以后几章所介绍的统计推断方法的工具,本身在实际中应用也很广泛。

第七章统计质量管理是工业企业全面质量管理中的重要组成部分,是第六章所介绍的方法的一个应用。对这方面不感兴趣的读者,可以不读这一章,基本上不影响后面几章的内容。但是据作者的经验,一个管理工作,即使不是质量管理工作者,如果他熟识统计质量管理的基本原理和方法,也许会应用到自己所面

对的并非产品质量管理的其他管理工作中。

第八章回归分析和相关分析，第九章抽样调查方法，第十章统计决策，这些都是在第四、五、六章基础上建立起来的管理中行之有效的统计方法。

本书可作为管理科学、经济管理、管理工程、管理信息等经济与管理类专业的教学参考书。作为教材，讲授全部内容，包括附录，约需 100学时；若省略附录及少部分较难的数学推演，约需讲授80学时。本书也可作为管理者、工程技术工作者的参考书。

1991.5

目 录

第一章 整理数据的几种简单方法	(1)
§1.1 数据分组方法.....	(1)
§1.2 图表示法.....	(8)
§1.3 数据中心趋向的度量.....	(12)
§1.4 数据离散性的度量.....	(17)
§1.5 利用数据分组表计算特征数.....	(18)
第二章 指数	(22)
§2.1 引言.....	(22)
§2.2 物价指数.....	(25)
§2.3 物量指数.....	(31)
§2.4 链型指数.....	(32)
§2.5 常见指数实例.....	(34)
第三章 时间序列	(37)
§3.1 时间序列模型.....	(37)
§3.2 趋势分析及预测.....	(41)
§3.3 季节变化分析.....	(55)
第四章 概率	(70)
§4.1 随机事件.....	(70)
§4.2 概率.....	(73)
§4.3 独立事件的联合概率.....	(77)
§4.4 条件概率.....	(80)
§4.5 全概率公式和巴叶斯公式.....	(82)
第五章 随机变量	(85)
§5.1 基本概念.....	(85)

§5.2	随机变量的期望值	(91)
§5.3	随机变量函数的概率分布及其期望值	(94)
§5.4	随机变量的方差	(98)
§5.5	正态分布	(101)
§5.6	多元随机变量	(106)
§5.7	条件分布、独立性	(110)
§5.8	多元随机变量函数的分布	(114)
§5.9	多元随机变量函数的期望值	(120)
§5.10	二项分布	(124)
§5.11	泊松分布	(128)
§5.12	附录	(131)
第六章	概率分布特征数的统计推断	(141)
§6.1	引言	(141)
§6.2	几个重要统计量的概率分布	(145)
§6.3	点估计的性质、极大似然估计法	(150)
§6.4	正态总体期望值的推断 (I)	(157)
§6.5	正态总体方差的推断	(165)
§6.6	正态总体期望值的推断 (II)	(169)
§6.7	两个正态总体期望值的比较	(172)
§6.8	关于比率的统计推断	(178)
§6.9	附录	(184)
第七章	统计质量管理	(189)
§7.1	引言	(189)
§7.2	抽样检验方法	(189)
§7.3	计数标准型抽样方案	(190)
§7.4	几种非标准型计数抽样检验方案	(194)
§7.5	计量型抽样检验方案	(198)
§7.6	生产过程控制图原理	(202)
§7.7	$\bar{x}$ -R 控制图	(206)
§7.8	p 控制图和pn控制图	(213)

§7.9 附录.....	(217)
第八章 回归分析和相关分析.....	(219)
§8.1 一元线性回归模型.....	(219)
§8.2 最小二乘法.....	(224)
§8.3 相关分析.....	(227)
§8.4 预测与控制.....	(234)
§8.5 多元线性回归.....	(238)
§8.6 偏相关系数.....	(243)
§8.7 非线性回归.....	(249)
§8.8 残差分析及Durbin - Watson 检验.....	(253)
§8.9 附录.....	(259)
第九章 抽样调查方法.....	(266)
§9.1 引言.....	(266)
§9.2 随机数表.....	(268)
§9.3 简单随机抽样.....	(270)
§9.4 分群抽样法.....	(275)
§9.5 分层抽样法.....	(282)
§9.6 社会敏感问题随机选答调查方法.....	(287)
§9.7 附录.....	(293)
第十章 统计决策.....	(303)
§10.1 引言.....	(303)
§10.2 最大风险最小化准则.....	(307)
§10.3 巴叶斯决策准则.....	(312)
§10.4 增量分析法.....	(316)
§10.5 决策树分析法.....	(320)
§10.6 信息价值.....	(322)
附表 1—10.....	(328)
部分习题的解答或提示.....	(363)
参考文献.....	(373)

第一章 整理数据的几种简单方法

在分析处理经济与管理问题时，经常使用大量的数据。例如要了解全国农民在实行联产承包责任制之后生活改善的状况，就需要对大量的农户的收支情况作调查。再如电视机厂要了解某批显像管的质量，就要对相当多的显像管某些指标进行测试。今后我们称上述类型的调查或测试为“观察”；称所要考察研究的对象的全体为“总体”；总体中每个元素称为个体；观察可对每个个体进行，也可以只观察一部分个体。所观察的部分个体，称为“样本”。一个样本所含有的个体的数目，称为这个样本的“容量”。

观察的结果，常常是一些杂乱无章的数，称为原始数据，或简称为“数据”。管理者要从数据中提取有用的信息，用来反映所观察总体的某些特性，必须先对数据加以整理。根据对数据不同的使用目的，常用的几种整理数据的方法是分组法、图示法以及特征数计算法等。这三种类型的数据整理方法又各自包括许多具体的方法。本章着重介绍这些方法的基本要求和步骤。读者学习和使用这些方法，可以根据面对的实际问题自行改进。评价这类方法的好坏，主要是看是否简便、适用，并没有严格的理论上的要求。

§1.1 数据分组方法

数据的分布情况，即在各种大小区段所占的百分比的情况；能够反映出所考察的总体的性质。原始数据一般是杂乱无章

的。为了较清楚地看出数据的分布情况，最基本的方法是将数据按一定的规则分组。下面我们将结合例子来讨论数据分组的一般要求、原则和步骤。

例1.1.1 某装配车间要考察一批轴承的质量，从该批轴承中抽取56个，测量其内径，其值为（单位：毫米）：

74.026	73.992	74.031	74.015	74.014	73.988	73.996	73.970
73.985	74.005	73.992	73.994	73.994	74.015	74.004	73.995
74.003	74.009	74.007	73.990	73.983	74.002	73.998	73.997
74.012	74.006	73.967	73.994	74.000	73.984	74.012	74.014
73.998	73.999	74.007	74.000	73.984	74.005	73.998	73.996
73.994	74.012	73.986	74.005	74.007	74.006	74.010	74.018
73.985	74.000	73.984	74.002	73.970	74.005	73.997	73.969

试通过对数据的分组整理，反映出数据的分布情况。

第1步：确定分组个数 k 。分组个数 k 是根据样本的容量 N 来确定，一般 N 越大，相应分组个数 k 也大。经验表明，当 N 在30至100时， k 取5到10比较适宜；当 N 在100至300时， k 取7到15比较适宜。分组太少，会损失较多的信息；分组过多，又不易反映出数据的概貌。

本例样本容量为56，确定分为6组。

第2步：找出数据中的最大值 L ，最小值 S ，并计算极差 R （定义极差 $R = L - S$ ）。

本例中， $L = 74.031$ ， $S = 73.967$ ， $R = 0.064$

第3步：确定组间距 h 。其公式为

$$h \approx \frac{R}{k}$$

一般应取 h 稍大于 $\frac{R}{k}$ ，以便在分组时能将最小值 S 包含在第一组，将最大值 L 包含在最后一组。为了简便，又要求 h 取位数尽可能少一些。

本例中, $\frac{R}{k} = 0.0106666$, 故取 $h = 0.011$.

第4步: 确定组界。关于组界的确定, 针对不同的实际问题, 有几种不同类型的方法, 后面我们还要讨论。这里, 我们要求所划分的各组的组界比原始数据多取一位数, 以使得在将原始数据分组时, 数据不至于落到组界上。

本例中, 原始数据的最小数为 73.967, 我们将第一组的左界取为 73.9665, 那么第一组的右界即为

$$73.9665 + h = 73.9665 + 0.011 = 73.9775$$

第 j 组的左、右界分别为

$$73.9665 + (j-1)h, \quad 73.9665 + jh$$

如表 1.1.1 的第 (2) 列所示。

表 1.1.1 数据 (频数、频率) 分组表

组号 (1)	组 界 (2)	组中值 (3)	频数 (4)	累计频数 (5)	频率 (%) (6)	累计频率 (%) (7)
1	73.9665 73.9775	73.972	4	4	7.1	7.1
2	73.9775 73.9885	73.983	8	12	14.3	21.4
3	73.9885 73.9995	73.994	16	28	28.6	50.0
4	73.9995 74.0105	74.005	18	46	32.1	82.1
5	74.0105 74.0215	74.016	8	54	14.3	96.4
6	74.0215 74.0325	74.027	2	56	3.6	100

第5步: 计算各组的频数、频率、累计频数、累计频率。

第4步确定了组界之后, 每个组就是一个确定的区间。称每个组, 即对应的区间, 所包含的原始数据的个数为该组的频数。

假定以 f_i 表示第 i 组的频数, 则称 $\frac{f_i}{N}$ 为该组的频率, 一般用百

分数表示。这里 N 表示原始数据的个数。一个组的频率，就是原始数据落在该组的比数。假定将第 j 组的频率记为 f_j^* ，则第 i 组的累计频数和累计频率分别定义为

$$F_i = f_1 + f_2 + \cdots + f_i \quad (i = 1, 2, \dots, k)$$

$$F_i^* = f_1^* + f_2^* + \cdots + f_i^* \quad (i = 1, 2, \dots, k)$$

本例中，各组对应的频数、频率、累计频数、累计频率如表 1.1.1 所示。

第 6 步：计算各组的组中值。假定以 a_i 和 b_i 分别表示第 i 组的左界和右界，则称

$$C_i = \frac{a_i + b_i}{2}$$

为第 i 组的组中值。在处理实际问题时，常用组中值作为该组的代表值。

本例中，各组的组中值如表 1.1.1 中第 (3) 列所示。

第 7 步：将第 4, 5, 6 步所得结果列表。

本例列表如表 1.1.1。

例 1.1.1 中，原始数据为 56 个，利用上述分组方法分为 6 个组。观察各组组界及各组的频率，可以清楚地看出数据的分布情况。

下面我们进一步讨论分组法中的几个问题。

数据分组方法是一种经验性整理数据的方法。要根据实际情况灵活地选取、调整分组法中的几个参数，如分组个数 k 、组间距 h 、各组的组界等。

在例 1.1.1 中，为了不使原始数据落到组界上，我们在确定组界时多取了一位小数。有时为了简便，也可以使用与原始数据有相同精确位数的组界，将各个组表示为形如 (a, b) 或形如 $[a, b)$ 的半开半闭的区间。特别对于那些只可能取整数的量，这种确定组界的方法是合适的。例如为考察某家用电器门市部日销售彩色电视机台数，现有 60 天的销售记录，其中日销售最少为 2 台，最多为 19 台。取组间距为 3，共分 6 组，其分组区间表示为

$[2, 5)$, $[5, 8)$, $[8, 11)$, $[11, 14)$, $[14, 17)$, $[17, 20)$, 这是左闭右开形区间。区间包含其左界, 而不包含其右界, 所以60个数据中的每一个必有确定的组别归属。类似地, 也可以表示为左开右闭形区间。

在分组时, 每个组的组间距也可以取不同的值。例如上面提到的日销售彩色电视机台数的例子, 若销售12台以上的天数是很少的, 而销售8台左右的天数又特别多, 根据实际情况, 我们也可以将数据分组为 $[2, 5)$, $[5, 7)$, $[7, 8)$, $[8, 9)$, $[9, 12)$, $[12, 20)$ 。

例1.1.2 1988年对山东省农户人均纯收入作了抽样调查, 调查户数为4200户, 其分组表如下表所示。

表1.1.2 山东省农户人均纯收入分组表 (单位: 元)

组号	组 界		频数	累计频数	频率 (%)	累计频率 (%)
1	100以下		15	15	0.36	0.36
2	100	150	57	72	1.36	1.72
3	150	200	86	158	2.05	3.77
4	200	300	369	527	8.79	12.56
5	300	400	606	1133	14.43	26.99
6	400	500	674	1807	16.05	43.04
7	500	600	710	2517	16.90	59.94
8	600	800	865	3382	20.59	80.53
9	800	1000	411	3793	9.78	90.31
10	1000	1500	327	4120	7.78	98.09
11	1500	2000	57	4177	1.36	99.45
12	2000以上		23	4200	0.55	100

表1.1.2就是一个不等组间距分组表实例。特别是第一组和最末一组，各有一端是无界的，也称为“开口”的。

上面所介绍的分组法是按事物的数量标志进行的。分组法也可按事物性质进行。

例1.1.3 消防部门为了分析、比较火灾的成因，以便加强消防管理，将火灾按引起的原因分组，见表1.1.3。

表1.1.3 1985年全国火因分组表

组号	火 因	频数	频率(%)	指数*
A	吸烟	3161	9.0	81.3
B	小孩玩火	3204	9.4	84.7
C	放火	1770	5.1	45.5
D	生活用火不慎	12919	36.9	322.2
E	违反安全操作规定	3541	10.1	91.1
F	违反电器安装使用规定	5214	14.9	134.1
G	自燃	703	2.0	18.1
H	其他	1944	5.6	50.0
I	未明	2450	7.6	63.0
	合计	34996	100	900

*参考§2.1

通过表1.1.3,人们可以清楚地看出火因的分布情况,例如生活用火不慎是引起火灾的很重要的因素,它所引起的火灾占全部火灾起数的36.9%。

练习题1.1.1 为研究男女婚配年龄问题,现有重庆市某两个街道的统计资料如表1.1.4所示。(资料取自《数理统计与管理》1986年第4期。)试根据表1.1.4的数据,整理成一个夫妻年龄差的分组表。

表1.1.4 重庆市解放碑街、沙坪坝街1979年以来结婚登记数

男婚 齡		20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40
女婚 齡		0	0	10	14	18	15	12	6	8	4	2	2	0	0	0	1	0	0	0	0	0
20		0	0	10	14	18	15	12	6	8	4	2	2	0	0	0	1	0	0	0	0	0
21		0	0	14	23	19	37	32	25	5	4	2	0	0	1	1	0	0	1	0	0	0
22		0	0	17	24	30	45	51	36	17	10	3	0	0	1	0	0	0	0	0	0	0
23		0	0	12	37	75	99	102	74	71	32	9	4	2	0	0	1	1	0	2	0	0
24		0	0	7	24	106	124	176	116	93	52	31	12	3	1	0	2	1	0	1	0	0
25		0	0	4	11	43	115	172	146	91	54	34	13	12	4	1	1	1	0	0	0	2
26		0	0	0	2	8	43	142	130	114	59	35	16	16	2	5	1	1	0	0	0	1
27		0	0	0	2	9	14	40	110	91	56	32	14	13	5	4	3	1	1	2	0	0
28		0	0	0	1	1	4	11	20	49	40	28	17	8	8	5	4	4	1	2	0	0
29		0	0	0	0	0	1	4	5	8	24	26	12	5	4	4	2	2	0	2	1	0
30		0	0	0	0	0	0	1	3	1	6	9	9	7	6	5	3	3	2	2	1	0
31		0	0	0	0	0	0	0	0	2	1	5	3	2	1	2	3	3	2	2	1	0
32		0	0	0	0	0	0	0	0	2	0	1	0	6	0	2	4	0	0	1	0	0
33		0	0	0	0	0	1	0	0	0	0	0	0	1	1	1	0	1	0	1	0	0

§1.2 图表示法

原始数据经过分组整理之后,容易看出其分布情况。如果用图形将数据分组表中所列的数据分布信息表示出来,会更直观,更便于应用。用来表示经济与管理数据性质的图形有很多类型,如条形图,直方图,面积图,多边形图(折线图)等。本节主要介绍几种类型的直方图和多边形图(折线图)。

(一) 频数直方图

在平面直角坐标系 oxy 上,以 ox 轴表示观察数据,将分组表中的组中值(或组界)标在 ox 轴上;在 oy 轴上表示出相应的频数。相应每一个组画一个矩形,其高为频数,底边在 ox 轴上,两端对应该组的左、右界。这种图形称为频数直方图。例如分组表1.1.1对应的频数直方图如图1.2.1所示。

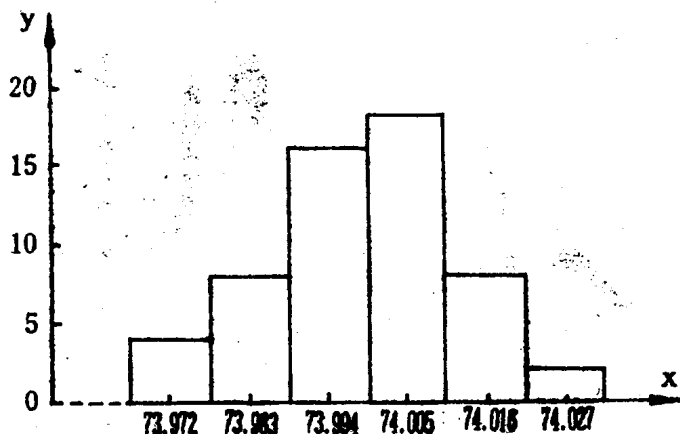


图1.2.1 轴承内径数据频数直方图

(二) 频率直方图

和频数直方图画法类似,只要以频率代替频数,便可画出频率直方图。