

方再根 杨述贤 编译

现代经营管理决策统计学

中国发明创造者基金会 中国预测研究会

现代经营管理决策统计学

方 再 根
杨 述 贤 编译

中国发明创造者基金会

中国预测研究会

1985·9

目 录

DN42/29

第一章 绪论	(1)
第二章 描述统计数据	(2)
第一节 总体和样本	(2)
第二节 频率分布	(4)
第三节 相对和累积频率分布	(9)
第四节 频率分布的一般型式	(12)
第三章 描述统计量度	(15)
第一节 统计、参量、决策	(15)
第二节 算术均值	(16)
第三节 中位数和众数	(18)
第四节 量度的变异性	(20)
第五节 比例	(27)
第四章 统计抽样研究	(28)
第一节 抽样的必要性	(28)
第二节 抽样研究的设计和实施	(29)
第三节 抽样偏见和误差	(31)
第四节 选择抽样	(33)
第五节 选择统计程序	(36)
第五章 概率	(37)
第一节 基本概念	(37)
第二节 事件及其间的关系	(40)
第三节 组合事件的概率	(42)
第四节 条件概率和联合概率	(46)
第五节 普遍的相乘法则、概率树和抽样	(48)
第六节 应用概率法则容易出现的错误	(52)
第七节 计算方法	(54)
第八节 应用贝阿斯理论修正概率	(58)
第六章 概率分布 期望值及抽样	(60)
第一节 概率分布	(60)
第二节 期望值和方差	(61)
第三节 均值抽样分布	(64)
第四节 二项式概率: 比例抽样分布	(68)
第五节 连续概率分布	(78)

第七章 正态分布	(81)
第一节 正态分布特性	(81)
第二节 求正态曲线下的面积	(84)
第三节 正态总体样本均值的抽样分布	(90)
第四节 总体非正态分布时 $\bar{X}$ 的抽样分布	(93)
第五节 P 的抽样分布和正态逼近	(96)
第六节 小总体的 $\bar{X}$ 抽样分布	(101)
第八章 统计估计	(103)
第一节 估计量与估计	(103)
第二节 大样本均值的区间估计	(105)
第三节 小样本 μ 的区间估计	(109)
第四节 总体比例的区间估计	(111)
第五节 确定必须的样本容量	(112)
第九章 假设检验	(117)
第一节 假设检验的基本概率	(117)
第二节 采用均值进行大样本检验	(121)
第三节 均值双边假设检验	(126)
第四节 利用小样本的均值检验	(127)
第五节 利用比例的假设检验	(129)
第六节 选择检验	(130)
第七节 假设检验方法的局限性	(133)
第十章 回归与相关	(134)
第一节 回归分析	(134)
第二节 最小二乘法	(140)
第三节 线性回归分析的假设和特性	(145)
第四节 使用回归直线时的预测和统计推断	(146)
第五节 相关分析	(150)
第十一章 多元回归和多元相关	(155)
第一节 包括三个变量的多元线性回归	(155)
第二节 用最小二乘法的多元回归	(156)
第三节 多个变量的回归——应用计算机	(162)
第四节 哑变量方法	(166)
第五节 多元相关	(170)
第十二章 时间序列分析和预测	(174)
第一节 时间序列及其组成	(174)
第二节 经典的时间序列模型	(175)
第三节 长期趋势的分析	(179)
第四节 用季节指数预测	(186)

第五节	周期和不规则起伏的辨识.....	(191)
第六节	指数平滑.....	(192)
第十三章	指数.....	(195)
第一节	价格指数.....	(195)
第二节	总价格指数.....	(196)
第三节	比价指数.....	(199)
第四节	消费品价格指数.....	(201)
第五节	用指数调节时间序列.....	(204)
第六节	数量指数.....	(205)
第七节	指数时间序列的移动和接合.....	(206)
第十四章	其他概率分布和拟合性检验.....	(209)
第一节	超几何分布.....	(209)
第二节	泊松分布.....	(214)
第三节	均匀分布.....	(218)
第四节	指数分布.....	(219)
第五节	拟合性检验.....	(224)

第一章 绪 论

统计学作为一门学科在本世纪发展很快，它已成为数学的一个重要分支。

本书采用的统计学定义是：为不确定条件下的决策提供数值依据的一批理论和方法。按这个定义，我们把统计学看作为一个独立的科学领域，就像物理学、历史学、生物学等一样。我们采用的统计学定义特别适用于主要从事经营管理和经济工作的读者，他们面临着许多不确定条件下的决策问题。

统计学可用来帮助人们制订方针政策，拟订长期规划，建立、验证和实现一些经济模型等。社会科学（特别是心理学和社会学），物理学、生命科学、遗传学等自然科学和工程技术各领域也都要应用统计学方法。虽然本书重点讨论的是统计学在经营管理和经济学方面的应用，而有些统计学书籍则是面向其他领域的，但是它们叙述的基本的统计学原理是一样的，其主要差别仅在于选用的例子和所强调的具体方法不同而已。

以前，统计学主要用来描述数据，称之为描述统计学。与之不同，现代统计学主要是通过研究局部的情况（称为样本）来推断整体情况（称为总体），这种统计学称为推断统计学。

我们还可将统计学分为演绎统计学和归纳统计学。用概率来确定获得特殊样本结果的可能性的统计学称为演绎统计学，它从已知的总体特性来研究可能的样本特性。归纳统计学从样本特性推断未知的总体特性。演绎统计学和归纳统计学完全是互为补充的。在我们学习如何从样本推断总体之前必须先学习如何取得样本，我们只有先学习演绎统计学才能去了解归纳统计学。

本书第二、三章讨论叙述统计学，第五、六、七章讨论演绎统计学，第八至第十四章讨论归纳统计学。

本书通过列举各种实际问题的例子，从直观或观念上讲解统计学的理论和方法。凡具有中等以上文化水平的读者都可读懂全书。

本书主要参考1981年美国出版的L. L. Lapin所著《Statistics for Modern Business Decisions》（第三版）进行编译。考虑到本书篇幅和读者对象，我们未编入原书中的一些章节和附录，读者可从原书参阅有关内容。本书第一章和第十一章至第十四章由方再根编译，第二章至第十章由杨述贤编译，杨念祖、杨大湘、方勤等参加了翻译工作。

第二章 描述统计数据

统计数据的整理和显示是描述统计的重要基础。单独的原始数字不能提供可以作出结论的图象。例如，我们可以相信医生通常有较高的收入；年青人的所得一般要比年长者少。但是，即便我们拥有国内税务局的全部税收资料，在对原始数据未作整理、筛选、分类时，我们也不可能确定内科医生所得是否比牙医高，个别律师所得是否比最好的医生多。只有对数据进行组织整理，我们才能获得用于事业管理的知识。大量数据是作出判断结论的原始根据，但对数据未作统计加工时，要想提出反映事物内在规律的见解是不可能的。大量数据可以认为是杂乱无章的。统计学并非只是收集大量数据，把这些数据用表格或图形表示出来。统计学也要研究实验的设计，抽样检查的设计，数据简化和处理。统计学是提供决策的工具，它是一门应用十分广泛的科学；自然科学和社会科学都用得上的一门科学。

第一节 总体和样本

统计学的基本单元是单个数据点，我们称它为观察值。因为它表示我们实际看到了什么。某一观察值可以是某一物理量度（重量或高度），对问题的一种回答（是或否），或一种分类（合格或不合格）。与具体决策有关的观察值构成总体。

定义：统计总体是与确定特性有关的一切可能观察值的集体。

总体含有一切可能观察值，其中有的是真的，有的是假的。与此相比样本只含有一些观察值。称某一问题的研究对象的全体为总体。

定义：样本是仅代表总体的一部分观察值的集合。

为了对总体的分布规律和数字特性进行研究，必须对总体进行抽样观察。从总体中，抽取若干个个体，这个若干个个体就称为一个容量（大小）为若干个的样本。由于每次抽取的数值可能不同，表现出一次抽样结果的具体数值称为样本的一个观察值。

不论是总体或样本的观察值，其集合均取决于分析目的。例如：某医学学会想要比较医生费用和医院行政其他部门成本。这里是两种总体：医生费用和医疗服务成本的比较。研究题目确定是否包括全国、一州或一县的费用和成本。

如果服务成本总体包括某特定日历年度内所有病愈病人时，对于某一单独患者病愈的收费就是一个样本。在同一年内，某一单个医院对病人的成本也是同一总体的样本。

一）基本单元

统计总体是由一些与具有个体某些特性有关的观察值所组成，而不是个体本身或个体项目。例如：某公司为了修改公司雇员退休计划，必须了解全体雇员年龄状况。由于年龄有许多变化，这就要搜集雇员个人档案和每个雇员的出生日期。这些确定的数字，就构成一个年龄特征观察值。全部雇员的所有年龄特征观察值，就构成公司雇员年龄总体。而雇员本身被称为总体的基本单元。

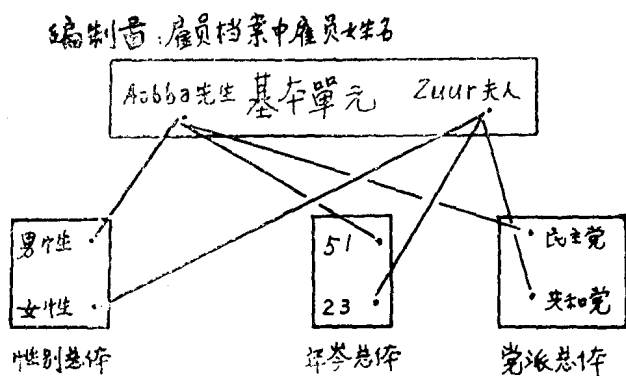


图2—1 说明由一个基本单元得到若干不同总体

二) 编制图

我们由所有雇员档案中得到的雇员名单，作为构成编制图的各种单元源。

在统计研究中，编制图是很重要的。因为用它来定义总体。假设采取投票人为样本，投票人对候选人是有偏好的。依此来预测选举结果。这种相关总体称为目标总体。以投票人为样本时，编制图只是记录投票人的名册。列入编制图中基本单元与目标总体中的单元不同，因为所记录的投票人有的不参予投票。所记录的人并投了票的，才应计算为投票人。由投票人构成的总体有些人是没有实际意义的，必须强调是参予投票的人才是有用的。样本必须由参予投票的人组成。

三) 定性、定量总体

以相关特性型式相区别的总体，基本上有两大类。能以数字表示特性的总体称为定量总体，如身高、重量、成本或收入等。特性是非数字的如性别、种族、婚姻状况、职业或专业等，其总体就称为定性总体。在统计学的描述中，我们将单独平行地讨论这两类总体，对定量总体用数字的算术运算可以完成。例如：我们计算30岁人所组成的集合的平均身高。但对定性总体的综述，必须使用另外的方法。

我们认为定性特征的观察值是总体的一种属性：如对于婚姻状况，即可由单身，已婚，离婚，丧偶等所表示。定量观察值为收入则是一个数值或一个变量如11.928，23…。表2—1给出各种定性和定量总体观察值示例。

表2—1 不同定量和定性总体的可能观察值示例

基本单元	定量总体观察值		
	相关特性	量度单位	可能变量
人	年龄	岁	21.3岁
微型电路	有缺陷的焊缝	数目	5条
轮胎	剩余的轮胎面	毫米	5毫米
账目平衡	数量	美元	5, 233.46美元
雇主	女性雇员	百分比	35%
通用股票	每股利润	美元	3.49美元
键盘穿孔机	错误	比例	0.02

灯泡	寿命	小时	581小时
食品罐头	成份重量	英两	15.3英两

定性总体观察值		
基本单元	相关特性	可能特性
人	性别	男的, 女的
保证措施	型式	合同, 通用股票, 优先股票
建筑	外部材料	砖, 木质, 铝
雇员	经验	合适的, 不合适的
电视机	质量	不合格, 合格
厂商	合法地位	股份, 合营, 专营
病人	状态	安全, 临危
学生	居住情况	住宿, 走读

第二节 频率分布

一) 确定数据有效图谱

表2—2给出了由100个学生组成的年龄样本。描述这一样本时,我们将如何着手进行呢?表2—2所列数据未经综合排列和整理加工。我们称其为原始数据。我们以一种有效方式将这些年龄数值分组处理。

表2—2

100个学生的年龄样本

年龄 (岁)									
20.9	33.4	18.7	24.2	22.1	18.9	21.9	20.5	21.9	37.3
57.2	25.3	24.6	29.0	26.3	19.1	48.7	23.5	23.1	28.6
21.3	22.4	22.3	20.0	30.3	31.7	34.3	28.5	36.1	32.6
33.7	19.6	18.7	24.3	27.1	20.7	22.2	19.2	26.5	27.4
22.8	51.3	44.4	22.9	20.6	32.8	27.3	23.5	23.8	22.4
18.1	23.9	20.8	41.5	20.4	21.3	19.3	24.2	22.3	23.1
22.9	21.3	29.7	25.6	33.7	24.2	24.5	21.2	21.5	25.8
21.5	21.5	27.0	19.9	29.2	25.3	26.4	22.7	27.9	22.0
23.3	28.1	24.8	19.6	23.7	26.3	30.1	29.7	24.8	24.7
23.5	22.9	26.0	25.2	23.6	21.0	30.9	21.7	28.3	22.1

一种实用法就是将年龄从18岁开始分为两种历年表。每个年龄均填在表内各列: 18.0—小于20.0, 20.0—小于22.0等等。我们称这种数值组合为等级区间。现在,我们将原始数据简化为每一组等级区间内的年龄数字,并将发现样本的某些特性。

我们首先按递增序列将等级区间编表,如图2—3第一列所示。而后,向下阅读原始数据表中的连续年龄。我们将一个标记符号相应地标在等级区间旁,以免两次计算,并校对初始

表格中数据旁的符号。标号完成后，计算符号数目，用以确定每组区间内观察值数目。最后结果表明在每个等级区间内所发生的年龄频率，称此频率为样本频率分布。在表2—3中，我们发现100个大学生统计值有10个值落在18.0—小于20区间内。我们称这10个数值为第一等

表 2-3 学生年龄样本频率分布

年 龄	标 记	人 数 (等级频率)
18.0 — 小于20.0	+++ ++	10
20.0 — 小于22.0	+++ +++ +	18
22.0 — 小于24.0	+++ ++ +++ +	23
24.0 — 小于26.0	+++ +++	14
26.0 — 小于28.0	+++ ++	10
28.0 — 小于30.0	+++	8
30.0 — 小于32.0	++	4
32.0 — 小于34.0	+	5
34.0 — 小于36.0		1
36.0 — 小于38.0		2
38.0 — 小于58.0	+	5
		总计100

级区间内的等级频率。每一个等级区间有两个极限。例如，第一个区间中下极限为18.0，上极限为20.0。我们用“小于”一词，严格区别于下一等级区间的极限20.0。这种规定，在于防止一个观察值在编排时，出现双值现象。例如19.99岁由第一等级区间给出，而20.01却落在第二区间。区间宽度是由它的上、下极限值而定。例如，第一个等级区间的宽度为 $20.0 - 18.0 = 2.0$ 。除最后一个区间外，所有区间宽度相等。正如我们所知，开始和最后区间的处理，往往不同。将原始数据作如此处理是有效的，这样就可构成一个总体频率分布表格。

二) 图表显示：频率曲线

在频率分布作图中，直观显示是一种有用的起点。图2—2中用图解法描述大学生年龄的频率分布。水平坐标代表年龄，且被分为两种年龄宽度的等级区间。与等级频率（每个区间的观察值数目）相应的不同高度柱状图形代表等级。于是，用垂直坐标就代表频率。这种频率分布描述称为频率曲线。

由于年龄23是在22.0—小于24.0岁这一区间，所对应的柱状图形是高为23。在34.0—小于36.0岁区间只有一个人，它的高度为1。注意年龄的标度是连续的。

注意最高等级区间（38.0—小于58.0）其宽为20，是标准区间的10倍。由于在38.0—58.0内仅有五个年龄标度，在此区间内的每两年的一个观察值，其均值为 $\frac{1}{2}$ 。所以其柱状图形高为 $\frac{1}{2}$ 。经过处理的两个年宽作为一个标准单位。任一柱状图形内的面积（其高为宽的若干标准单位倍数）等于相应等级区间内的观察数目。于是，在第一个柱状图形内面积为 $10 \times 1 = 10$ 。而最后一个柱状图形面积为 $\frac{1}{2} \times 10 = 5$ 。

图表显示：频率折线和曲线

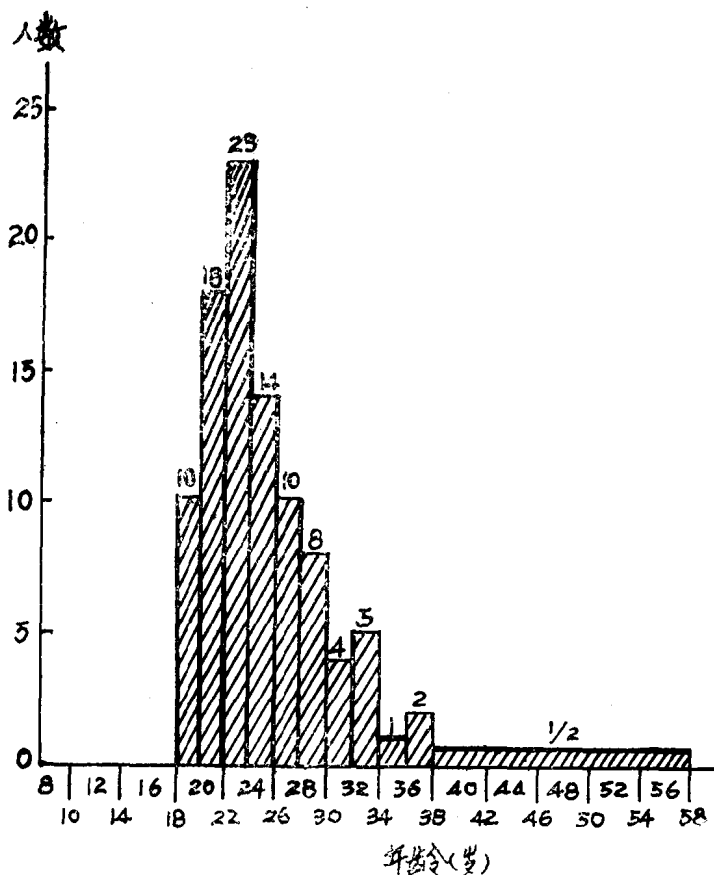


图2—2 大学生年龄频率分布图

一种可供采用的频率分布描述就是图2—3所给出的同一个大学生年龄样本的频率折线。在图2—3中，每个等级区间由位于该区间中间的圆点表示，且圆点高度为该等级区间的频率。每个圆点的垂直座标代表等级频率如第一个圆点所对应的等级频率为10。圆点的水平座标代表对应区间的中点。这样，第一个圆点的位置的确定：垂直座标为其等级频率10，水平座标为该区间中点 $(18.0 + 20.0) / 2 = 19.0$ 。所以，第一个圆点的座标为 $(19, 10)$ 。类似，第二个圆点的位置：其等级频率为18，该区间中点为21，所以该圆点的座标为 $(21, 18)$ 。这样，图2—3中某一圆点的位置：水平轴座标上由对应区间的中点值决定，垂直座标由相应的等级频率决定。最初和最后圆点的确定：因为该两点对应的等级频率为0，所以均位于水平轴上；其距原点水平距离由相邻圆点的水平距离可知为标准宽度一半的最低区间和最高区间，即x轴上17和19两点。联结上述所有圆点就得到频率折线。

当观察值很大时，样本频率折线就变得类似整个总体数据所得的图形。但是，这种总体图形，最初并不适用。频率分布曲线常常要描绘成如图2—4所示的型式。这种图形叫做整个总体频率分布光滑曲线图。

三) 叙述分析

把杂乱无章的大学生年龄原始数据变换为频率分布图型。由此图型可以了解两种事情：观察值组合围绕一个中心值如何分布，各观察值间的散布和偏离的程度。

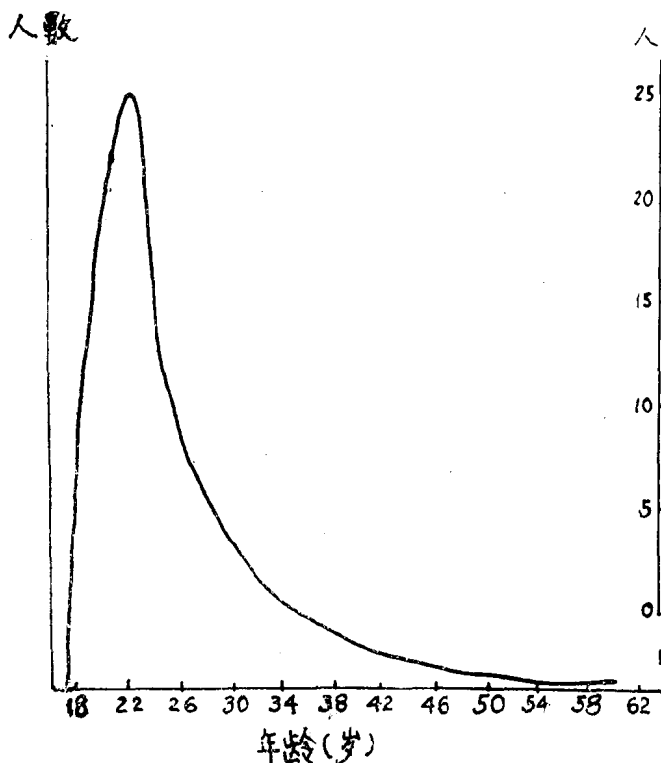


图2—3 大学生年龄样本频率折线

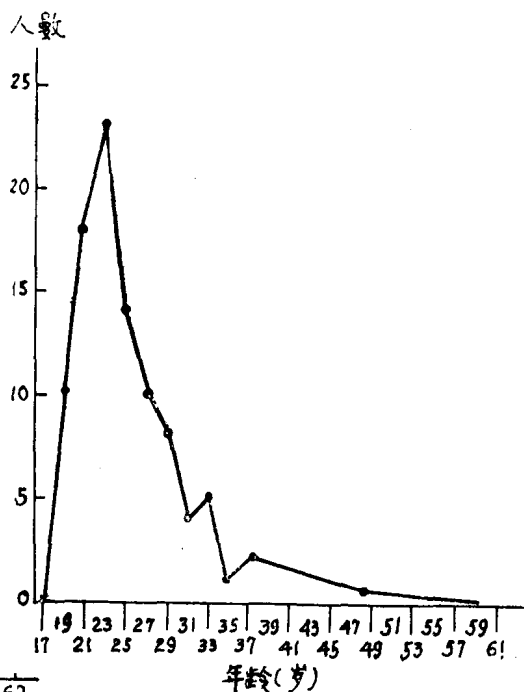


图2—4 整个总体频率分布光滑曲线

我们知道大学生小于18岁的极少，多数在28岁以下。最普通的年龄是在22—24岁之间，这一范围是比高中刚毕业考进大学的学生年龄要大，比大学生毕业年龄要小一些，一般大学毕业为22岁。

由叙述分析中，我们可以得到一个十分重要的概念：原始数据不是立即可用的。频率分布是叙述的基础。我们将在以后的章节中看到频率分布是概率理论的基础，它在一切统计推论过程中起着十分重要的作用。

四) 频率分布构成

频率分布将原始数据变换成为统计分析的一幅有意义图型。频率分布把许多变量归并为等级区间，但它并不能告诉我们在某一范围内观察值之间彼此的区别。只有原始数据才能说明这一点。这样，我们就保留有用的数据和丢弃不太重要的若干数据。

如何最恰当地将原始数据变换为等级区间？这一问题包含两个方面：应该采用多少区间？哪些数据应该包括在一个单独的范围之内？

等级区间的宽度应取为等宽度以便进行比较和提供更为简明扼要的数值估计。正如表2—3中年龄数据分布所示，将特大值归并为一个单独范围里，它将比频率大若干倍。在组成一个个人年收入频率分布时，大多数人所赚得的收入小于5万美元，而有些人的收入在5万到10万之间。非常富裕的人才赚得突出多的收入。一个具有等区间收入宽度为5000美元的频率分布曲线大概是绘制费力，阅读困难。最后的频率分布曲线应该和较低水平的频率山峰形状相似。以递增序列划分区间。通常用5到20个等级区间就可以了。确切的数目，则是一个人的

判断力的问题。当全部区间是等宽度时，下列准则可以用作确定区间宽度：

$$\text{宽度} = \frac{\text{最大数值} - \text{最小数值}}{\text{等级区间数目}}$$

表2—4

100辆大卡车耗油量原始数据

19.0	20.8	22.0	22.7	20.0	18.9	16.6	16.8	20.8	14.7
15.1	21.8	21.1	21.5	21.1	15.5	19.3	15.1	20.6	16.8
18.2	20.5	15.3	16.2	16.3	22.8	22.7	21.9	22.5	17.1
19.1	21.6	19.0	18.3	18.6	22.1	17.5	22.9	21.7	18.7
21.9	20.2	14.5	14.1	22.9	20.2	17.3	22.6	19.3	21.7
21.5	22.6	18.7	19.2	22.8	21.6	21.7	20.5	22.7	20.4
18.8	15.1	16.5	20.5	19.1	17.4	19.7	19.2	16.4	21.9
14.3	19.2	19.7	17.1	21.4	21.9	21.7	19.2	23.9	19.6
20.9	18.5	20.2	18.2	20.2	22.4	20.4	21.6	21.3	22.4
20.5	18.1	20.7	21.3	16.9	20.3	23.9	18.8	21.1	21.9

举例说明：如表2—4中给出一个汽车运输队100辆汽车耗油量以每哩加仑计算的观察值。假设我们希望构成五个等宽度的等级区间。我们以最大值和最小值之差 $23.9 - 14.1 = 9.8$ 并除以5则得到所需要的宽度为1.96mpg。为了简化起见，可将1.96化为整数2.0作宽度，将14.1化为14.0作为第一个等级区间的最高极限。由表2—5和图2—5(a)提供这一频率分布。

表2—5

100辆大卡车耗油量的频率分布五个等级区间

每加仑英里	卡车数目
14.0—16.0以下	9
16.0—18.0以下	13
18.0—20.0以下	24
20.0—22.0以下	38
22.0—24.0以下	16
	总计100

图2—5中还包含有用10个和20个等级区间，对同一耗油量数据作频率曲线为图2—5(b)和图2—5(c)所示。值得注意的是随着等级区间数目的增加，曲线外形就变得更为凹凸不平。由五个和十个区间的频率曲线可以看出一种数据累积现象。但是，在区间选取20个时，明显的“锯齿”现象发生了。区间数目的多少，应该通过多次试验来决定。

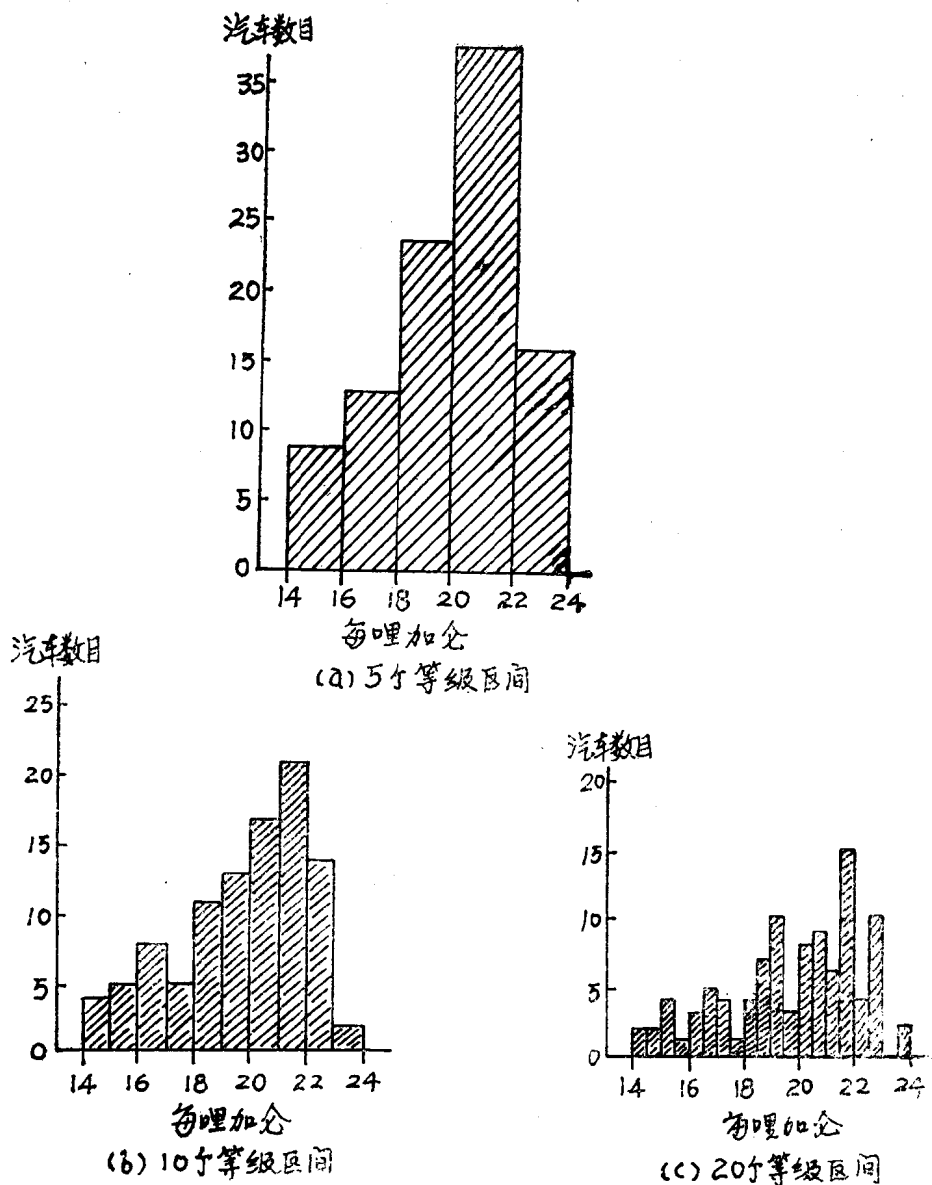


图2—5 使用三种等级区间100辆卡车耗油量频率图

第三节 相对和累积频率分布

基本频率分布的两种有用的推广是相对和累积分布。

一) 相对频率分布

相对频率指的是在某类内观察值数目与观察值总数目的比。相对频率可以用于确定定量和定性的两种数据，它是比较大小不同的各个相似组的基础。例如，比较两种地区的家庭收入。一地居住的人口稀少如在内华达 (Nevada) 州，另地人口密集如加利福尼亚 (California) 州。如果只讲在Nevada州6千家的家庭收入达到了California州25万家相同收入是没有意义的。但是如果说在Nevada州1%家和California州0.08%家达到了上述收入水平

时，我们就可以在这两个州的收入中进行一次逼真比较。我们必须用比例（百分数）而非全频率，进行比较，在每个州中看哪个州的收入达到上述收入水平。

总体相对频率的计算是由每个范畴内的观察值数目被观察值总数目相除。已知频率分布，相对频率的计算仅是一种简单的算术运算。表2—6举例说明百货公司应收账款使用年限数据的等级区间（多于180天的账目就没有收集）小于30天的老账目（其等级区间为0—小于30天）的相对频率为 $532/1615 = 0.330$ ，即它是老账目的33%。其余不同年限的相对频率计算

表2—6 百货公司应收账款使用年限相对频率分布计算

(1)	(2)	(3)
年限 (天数)	账目数 (频率)	相对频率
0—30以下	532	0.330
30—60以下	317	0.196
60—90以下	285	0.176
90—120以下	176	0.109
120—150以下	158	0.098
150—180以下	147	0.091
共计1615		1.000

也是用栏（2）中的等级频率被总账目数除。所以相对频率之和必须为1。由栏（1）和栏（3）的相应数据组成相对频率分布。图2—6给出这种分布的频率曲线。相对频率分布和绝对频率分布的曲线有相同的形状。只是垂直座标的比例变化而已。

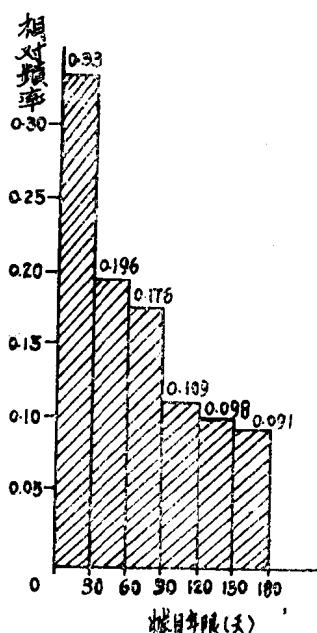


图2—6 百货公司应收账款
使用年限相对频率分布

当用绝对频率分析无效时，则用相对频率分析。公司的赊销量随着季节波动，但同一季节的图形一般是逐年保持的。如果顾客信贷价值保持在和以往大致相同水平时，就可得到应收账款使用年限相对频率分布。如果增长是稳定的，相对频率分布在连续两年的历月期间内不应有明显的不同。不同身份的顾客和积累过程变化时将引起相对频率变化。

相对频率也可用作两种定性总体的比较。

表2—7列出两所大学大学生性别频率分布。在学校间举行比赛时，一所大型大学有6000人男生参加，而在一所学院里只有400人男生参加。在表2—8中给出相对频率分布。就相对频率而言，在大学里参加比赛的男生比例则小于学院里男生比例。在大学的男生比例为0.6，而在学院里男生比例是0.8。

表2—7

大 学 生 性 别 频 率 分 布

大 学		学 院	
性 别	频 率	性 别	频 率
男	6000	男	400
女	4000	女	100
共计10000		500	

表2—8

大 学 生 性 别 相 对 频 率 分 布

大 学		学 院	
性 别	频 率	性 别	频 率
男	0.64	男	0.80
女	0.40	女	0.20
共计1.00		1.00	

二) 累积频率分布

只有当观察值是以数字表示时,累积频率才是逐个较大等级区间的频率和。例如,我们也许在10000人的总体中发现身高低于6呎的有8439人。另一种表示法,是低于6呎的人占84.39%。累积频率,尤其是以相对百分比或相对比例表示时,它对整个能够提供有效的描述。例如,我们以零售商店年销售量100台电视机为样本。销售单位以每100台来划分。在表2—9中给出零售商店年销售量100台电视机的示例。

累积频率是由100—200台以下零售商店数(频率)加上200—300台以下频率即得到200—300台以下的累积频率。亦即 $0 + 5 = 5$ 。某一等级区间的累积频率表示在此区间内提取销量水平的总和。由表2—9中的栏(1),栏(3)构成了电视机销售数据的累积频率分布。

表2—9

零售商店电视机销售的累积频率

(1)	(2)	(3)
销售单位(台)	零售商店数(频率)	累积频率
100—200以下	0	0
200—300以下	5	$0 + 5 = 5$
300—400以下	39	$5 + 39 = 44$
400—500以下	31	$44 + 31 = 75$
500—600以下	16	$75 + 16 = 91$
600—700以下	6	$91 + 6 = 97$
700—800以下	3	$93 + 3 = 100$
	100	

在图2—7中以图线形式给出了电视机销售数据的累积频率分布。图2—7的横座标代表电视机的销售量(台数),纵座标代表累积频率。例如表2—9中栏(1)中,400—500台以下时所对应的栏(3)中累积频率为75。于是,在图2—7中对应的是圆点座标为(500,

75)。利用连续相对频率相加就可得到累积频率。最大可能的累积相对频率为1。

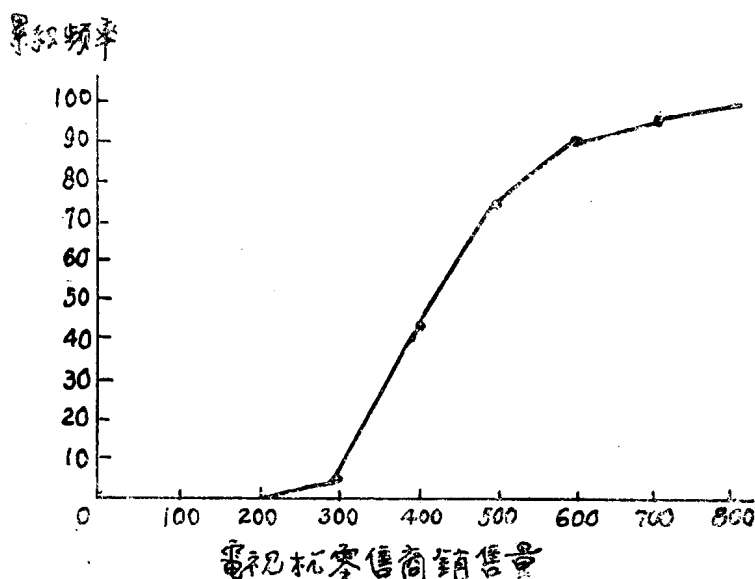


图2—7 电视机零售量累积频率曲线

第四节 频率分布的一般型式

由于对各种型式的总体样本的研究，使得统计方法得到发展。将各种总体归纳为不同类，就大大地减少了分析工作。定量总体可以由若干种分布型式加以描述。这样总体就可分为几个著名的分布，如正态分布，二项分布等。虽然描述这些分布的数学方程可能很复杂，但样本频率分布曲线的形状，一般可以表征总体分布的形状。

在下列图表中，给出了几幅通用的频率曲线。与这些曲线相应的样本频率是有代表性的。例如，在图2—8中的曲线，它代表在同一生产批量中，一时粗的钢筋棒，其直径的相对频率分布。应当指出这种图形是很有代表性的。许多频率分布是属于这种类型。图2—8中所给出的左图形对于区间0.995—1.005是完全对称的。远离这区间的柱状线条愈来愈短，约在0.91或1.09时处逐渐停止，柱状线条高度为零。与此柱状线条图所对应的平滑频率曲线是一条钟形曲线称为正态分布曲线。在本书的第七章中专门讨论正态分布。

图2—9表示一个由表2—9所给出的电视机销售数据频率曲线。图中数据柱状线条向右倾斜，这是因为某些商店的销售水平非常高所造成。与此频率曲线相对应的曲线向右结尾比左端大。这种图形与一种倾斜分布相应，图中有一些小观察值和比较大的观察值落在边远范围。

图2—9中给出了另一种倾斜分布型式。它和图2—10刚好相反。它的左端结尾比右端为大。这是一幅汽油每哩加仑耗油量频率曲线。在此情况，比较少的数值散布在14—18每哩加仑之间，稀少的大数值落在单独等级区间23.0—24.0内。

表2—6所给出的百货公司应收账款使用年限的频率分布绘制于图2—11中。这种型式的频率分布称为指数分布。其总体分布的光滑曲线形状与英文字母J的反向形状相当。指数分