Harvard Business School

Master of Business Administration

哈佛商学院MBA教程

下

真正地读一两本哈佛工商管理硕士教程
胜过听一列车的哈佛神话

人们迷信哈佛MBA
我无可奈何　但我劝他们
还是坐下来真正看一点他们所学的东西

经济日报出版社

本书工作人员

策　划:德　仁　张　弘

主　编:圣　丁

副主编:丁春生　张同全　欧阳云　叶军芳

编　委:丁春生　张同全　欧阳云　叶军芳　秦　风
　　　　刘　易　王仁生　张　征　刘晓和

第十章

统计学

第一节 统计学的研究方法

统计学研究对象的性质决定着统计学的研究方法。以下阐述统计学上所特有的基本方法:大量观察法、综合指标法、归纳推断法。至于一般的统计方法,则分别在其它章节中论述。由于统计学所研究的是特定总体数量关系的方法论,这些方法的数学依据是大数定律,所以本节中也讲述大数定律的意义和作用。

一、大量观察法

所谓大量观察法,是指对所研究事物的全部或足够数量进行观察的方法。社会经济现象是受各种因素相互影响的结果。在社会现象的总体中,个别单位往往受偶然因素影响,如果只选择一部分单位进行观察,是不能代表总体的一般特征的,必须观察事物的全部或足够数量单位并加以综合分析,这样可使事物中非本质的偶然因素的影响互相抵消或削弱,社会现象的一般特征就能显示出来。这是由于统计研究对象的多样性和复杂性所决定的。

政治算术学派的苏斯密尔希,被称为大量观察法的倡导者。他从人口规律的研究中得出结论:"事实若多一分,人事现象的规律则多发现一分;事实若少一分,人事现象的规律则少发现一分。""因此,不能用太少的事实,要尽可能地多,而且更好的是要尽可能包含更多的年份。"

社会统计学派梅尔认为,统计学研究的是社会总体而不是个别的社会现象,由于社会现象的复杂性和总体性,必须对总体

进行大量观察和分析，研究其内在联系，方能反映社会现象的规律。以大量观察法作为统计研究的方法是由梅尔全部完成的，并成为 19 世纪末到 20 世纪初支配德国统计学界的主要理论体系，统计调查运用大量观察法，可以对总体的所有单位进行全面调查。大量观察法并不排斥对个别单位的典型调查，它可以同典型调查结合起来，加深对社会现象的认识。

二、综合指标法

任何统计对象的具体项目，都是以统计指标表示的。统计所要了解的，不是个别事物的数值，而是将个别事物的数值综合汇总而成总体数值，这个数值是通过综合指标反映的。综合指标表示的具体现象的总体数量关系，它包括指标名称和指标数值。社会经济现象的统计指标是根据政治经济学以及其他有关的社会科学而确定的。

常用的综合指标有总量指标、相对指标、平均指标等，这些综合指标概括地描述了总体各单位变量分布的综合数量特征。综合指标法是统计分析的基本方法，其他各种统计分析方法都以它作为基础。如时间数列法、指数法、抽样法、相关法等都离不开综合指标的对比分析。

三、归纳推断法

所谓归纳法是从个别到一般的推理方法，是统计研究中常用的方法。上面所谈的综合指标法将个别现象的数值综合汇总成总体数值，概括反映了总体一般的数量特征，所采用的方法就是归纳法。

在研究社会经济现象的总体数量关系时，我们需要了解的总体对象范围常常是很大的，有时甚至是无限的，以致只能从中观察宣传部分单位进行计算和分析，根据结果来推论总体。例如，正在流水线上大规模生产的产品零件，需要及时了解它们的质量，只能抽取其中的部分产品检验，借以推断这一批产品质量的好坏，并以一定的置信标准来推论所作结论的可靠程度。

这种根据样本数据来判断总体数量特征的归纳推理的方法称为统计推断法。统计推断是现代统计学的基本方法，这种方法既可以用于对总体参数的估计，也可以用作对总体的某些假设检验。

四、大数定律

大数定律最早由瑞士数学家雅各布·伯努里提出，后由凯特勒完成了大数定律和概率论同统计的结合。他从莫阿弗尔，拉普勒斯等人所证实的大数定律的数学定理出发，又吸收了格朗特、苏斯密尔希等人对社会经济现象所提出的大数法则思想，从而首次在社会科学的范畴内提出他的大数定律的思想，并把统计理论建立在大数定律的基础上。在他的代表作《论人类》中，他通过大量统计资料的计算论证，社会生活现象并非偶然结合。他说根据现象的大量数据所产生的、概括的平均数，反映着人们所研究的现象的典型面貌及其发展规律。

大数定律是随机现象出现的基本规律，也是随机现象大量重复中出现的必然规律。总体中所包含的个体存在着共同的规律性，这种规律性只有在大量观察中才能显示出来。大数定律的本质意义，在于经过大量观察，把个别的、偶然的差异性互相抵消，而必然的、集体的规律性便显示出来。

社会经济现象也具有随机性，因为社会经济现象是受多种原因影响的，而这些原因的发生有许多的偶然因素，有些是必然因素，这两种因素总是交错结合在一起。分析社会经济现象必须对总体中全部或足够数量的单位进行观察，通过平均数将偶然的随机因素抵消，从而显示出现象的典型水平。

第二节　如何开展一个项目的研究

管理者怎样才能获得能够做出最优决策的信息呢？下面就开始我们的学习。首先介绍一下统计学的应用范围：

财政收入预测

竞争威胁的判断

消费者态度调查

广告效用的评价

价格政策的制订

质量控制标准评价

管理者收集的信息来自各种渠道。这些信息要经过整理和分类，才能作为决策的依据。一些信息是用来做主要研究工作的；另外一些则用于辅助研究工作。取得辅助研究的资料很容易而且费用不高。他们可以从以下的途径获得：

市场调查公司

广告代理机构

促销公司

公关部门

贸易协会

行业协会

政府机构

半官方机构

专业杂志和出版物

商业杂志和出版物

报纸的研究部门

杂志的研究部门

邮寄公司

科学研究机构

主要研究工作迄今为止还不具有创造性，可以说只有固定的模式。其资料常常可利用上述途径得到，然后就能进行研究了。

一、分析数据

整理检查完原始数据后，可以用定量的方法分析数据。特别是描述变量的概率及相关关系时，用定量方法分析数据所得到的结果，能使结论的表述更加简练。例如：90%的被访者都认为，自已更喜欢新品牌的某产品，而不是 X 牌的旧产品。

二、频率分布

在进行数据分析时，人们感兴趣的是数据(观察值)集中或者离散、相似或者差异的程度。数据的波动程度是指数据的离散度，可以用图形来表示，即图形分布。

图形分布大致可分为三类：

(一)对称分布(正态分布)

对称分布的图形如钟形(如图 10—1)。大多数观察值落在图形的中部，其他的均匀分布在两边。例如人的身材，过于矮小和过于高大的人，在整个人群中所占的比例极小；较为矮小和较高大的人占较大的比例；而中等身材的人占最大的比例。所以，

我们说，人在身材上的比例分布基本上属于对称分布。

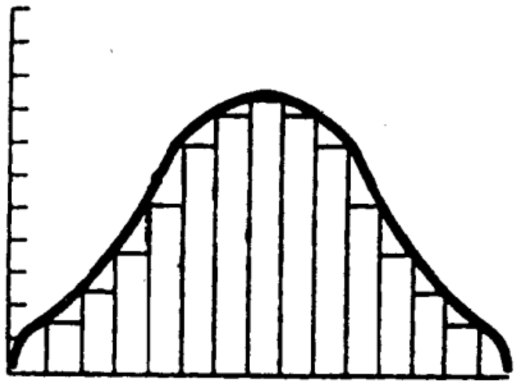

图 10—1　对称分布

(二)偏斜分布

偏斜分布的图形如图 10—2 所示。它是指大多数观察值分布在一边(或说聚集在一边)，而较少数处于另一边。它又有两种形态：一种是如图 10—2(a)所示的向下偏斜；另一种是图上出售的各种产品的数量，取决于他希望得到的价格。需求与供给的分析方法，是把这两种力分别开来，轮流分析，然后再把它们合到一起，说明价格和产出。

A 先生要买多少某种消费物品？第一，这取决于他的个人偏好。有的人喜欢菠菜，而别的人却不喜欢，因而购买情形不同。第二，这与 A 先生的收入大有关系。个人收入愈高，他购买各种物品愈多。

第三，物品的价格有一种效应。总的说，价格越低，购买量越大。最后，有关物品的价格也有关系。可能联系到另一种物品，有的是因为可以代有，或是因为正常情况下一块儿使用(用作补充)。奶油和植物油是可以相互代用的物品。奶油价格高了，购

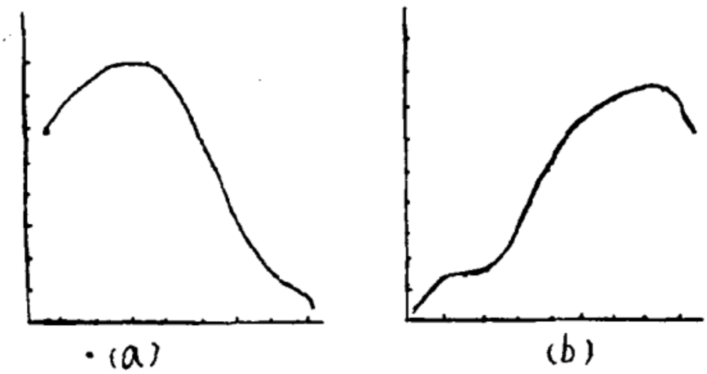

图 10—2　偏斜分布

买植物油就会多起来。汽油和汽车是相互补充，汽车价格低些，汽油的购买量就大起来。

充分地说明消费者的行为，要求在所有这些方面进行探究。但是，在本章，探究重点放在需求方面：购买物品的数量与该物品价格的关系。在分析时假定购买者的收入、口味，一切有关联的产品的价格，都是既定的，都是常数。

这一切都肯定了，那么关于 A 先生对某一特定物品，比如说咖啡的需求，我们能说什么呢？在这个意义上，需求是什么呢？怎么叙述它呢？

需求是按各种价格要购买的数量单(或表)。在经济学中，“需求”这术语，说的总是一个表。它不是单一的数量。如果要侧重研究按某种特定价格购买的数量，它是所需求的量(需求量)。

要说明 A 先生对咖啡的需求，必须按咖啡的各种可能的价格，指明每种所需求的数量。例如，A 先生的需求可能象图 10—2(b)。如果价格是 2.5 美元，他就只买 1 磅。价格低些，他就多买些。如果价格降到 25 美分，他就每月买 10 磅。

10—2(b)所示的向上偏斜。例如,假定两个群体都以图 10—2 的横轴代表收入水平,收入水平从低到高排列,假设两群体中所拿最低收入相同,最高收入也相同。纵轴代表拿某种收入人数的话,那么就可以看出,图 10—2(b)所代表的群体;比(a)所代表的群体要富裕。

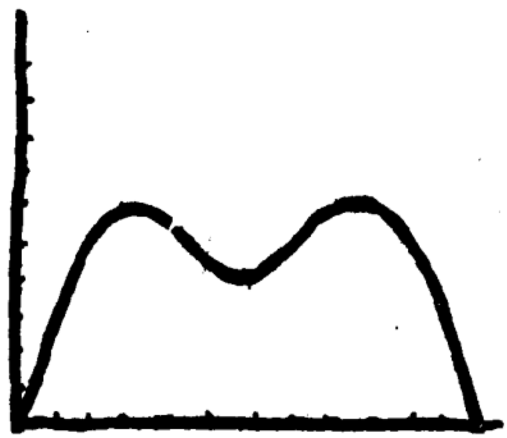

图 10—3　双峰分布

(三)双峰分布

这种图形的形状像驼峰,见图 10—3。它表明观察值形成为两个集中区域。例如某班级考试的成绩,得较低分数和较高分数的人都较多,只有少数人得中等分数,所以形成双峰分布。

第三节　集中趋势和离散趋势

一、集中趋势的测试

常用来表述数列集中趋势的测试有算术平均数、调和平均数、几何平均数、中位数和众数。这些测试在统计学中也为平均指标或平均数，可以用来反映标志值的典型水平和标志值分布的中心位置或集中趋势。

(一)算术平均数

算术平均数是平均指标中最重要的一种。一般不特别说明时所称的“平均数”就是指算术平均数。算术平均数的定义公式如下：

$$算术平均数=\frac{总体标志总量}{总体单位总量}$$

例：某企业工人月工资总额为 80,000 元，人数为 2,00 人，则该月工人的平均工资$\frac{80,000}{2,00}=400$ 元。

计算算术平均数时，标志总量和单位总量必须属于同一总体，分子和分母所包含的口径必须一致。否则，计算出来的平均数指示便失去科学性。利用以上定义式计算的平均数，可分为简单算术平均数和加权算术平均数两种。

1. 简单算术平均数

计算算术平均数在不具备总体的标志总量和单位总量资料时，要依据总体各个单位的具体资料来计算。将总体的各个单位标志值简单相加，然后除以单位个数，求出的平均标志值，叫做简单算术平均数。简单算术平均数的计算公式为：

$$\overline{X} = \frac{X_1 + X_2 + X_3 + \cdots + X_n}{n} = \frac{\sum_{i=1}^{n} X_i}{n}$$

式中：$\overline{X}$ 代表算术平均数；

X_i 代表第 i 个单位的标志值，i=1，2，3，…n；

n 代表总体单位数；

$\sum$ 代表总和。

例：企业车间中有 12 个工人，每个工人日产某种产品件数为：17，15，18，16，17，16，14，17，16，15，18，16，则某工厂生产班组的平均日产量用简单算术平均数计算如下：

$$\overline{X} = \frac{17+15+18+16+17+16+14+17+16+15+18+16}{12}$$

$$= \frac{195}{12} = 16.25(\text{件})$$

2. 加权算术平均数

当资料中被平均的变量值重复出现时，例如某个变量值 X 重复出现 f 次，按照简单平均法，就要对变量值 X 连加 f 次。为了简化计算，可以用 X 乘 f 来代替同一变量值 X 的连加 f 次。用这种方法计算的平均数就称为加权算术均数。其计算公式为：

$$\overline{X}=\frac{\sum_{i=1}^{n}X_if_i}{\sum_{i=1}^{n}f_i}$$

简写为：

$$\overline{X}=\frac{\sum Xf}{\sum f}$$

式中：X_i(或 X) 代表标志值；

f_i(或 f) 代表标志值 X_i(或 X)B 出现的次数；

n 代表组数；

$\sum_{i=1}^{n}X_if_i$(或$\sum Xf$)代表总量。

表 10—1　某车间工人数及日产量资料

工人按是零件分组(X)	工人数(f)	每个工人生产件数×工人数(Xf)
20	10	200
22	12	264
24	25	600
26	30	780
30	18	540
32	15	480
33	10	330
合　　计	120	3,194

根据上表资料，计算中权算术平均数为：

$$X=\frac{\sum Xf}{\sum f}=\frac{3,194}{120}=26.61(\text{件})$$

$\overline{X}=\frac{\sum Xf}{\sum f}$ 式表明，平均数的大小，不仅取决于总体各单位的标志值，而且受单位标志值出现的次数的影响。所以，统计学里把 $\overline{X}=\frac{\sum Xf}{\sum f}$ 式中的 f_i(或 f) 也称作为权数。权数可以分为绝对

数权数和相对数权数。记相对数权数 $W_i=\frac{f_i}{\sum Xf_i}$,则 $\overline{X}=\frac{\sum Xf}{\sum f}$ 式的加权平均数可以表述为:$\overline{X}=\sum X_iW_i$　其中:W_i 是 X_i 的权数(X_i 出现的相对次数),即它对平均的结果起权衡轻重的作用。

显然,如果各组次数完全相等,则 f 对各组标志值产生同等的影响,它不再起权衡轻重的作用,这时加权算术平均数就等于前述的简单算术平均数,所以可把简单算术平均数看作是加权算术平均数的一个特例。即,当各组的次数 f_i 相等时:

$$f_1=f_2\cdots\cdots=f_n=f。$$

则:

$$\overline{X}=\frac{\sum Xf}{\sum f}=\frac{f\sum X}{nf}=\frac{\sum X}{n}$$

如果你所掌握的资料不是单项数列资料,而是组距数列资料时,计算算术平均数的方法与上述方法基本相同。只是先要计算出各组的组中值$\frac{(下限+上限)}{2}$,以各组组中值代表该组标志值进行计算。

例:某企业工人每月工资分组资料如表(10—2):

表 10—2　某企业每月工资分组资料

月工资分组(元)	组中值(元)(X)	工人人数(f)	各组工人工资总额(元　Xf)
50～60	55	10	550
60～70	65	10	650
70～80	75	30	2,250
80～90	85	40	3,400
90～100	95	10	950
合　计	—	100	7,800

以各组的组中值为标志值代入中权算术平均数的公式得:

$$X=\frac{\sum Xf}{\sum f}=\frac{7,800}{100}=78(\text{元})$$

利用组中值计算算术平均数，是以假定各组内的标志值均匀分布为前提，而计算的结果同实际情况可能会有一些偏差，因此是平均数的近似值。

(二)调和平均数

调和平均数又称“倒数平均数”，它是根据各标志值的倒数来计算的平均数。具体地说，调和平均数就是各个标志值倒数的算术平均数的倒数。但计算结果并非是算术平均数的倒数。调和平均数应用并不广泛，统计工作中往往是把调和平均数的计算形式，作为算术数的变形来使用。

调和平均数也分简单调和平均数和加权调和平均数。如以 $\overline{X}_h$ 代表调和平均数，以 n 代表资料项数，则简单调和平均数的计算公式为：

$$\frac{1}{\overline{X}_H}=\frac{\frac{1}{X_1}+\frac{1}{X_2}+\cdots+\frac{1}{X_h}}{n}$$

即：$$X_H=\frac{n}{\frac{1}{X_1}+\frac{1}{X_2}+\cdots\frac{1}{X_h}}=\frac{n}{\sum\frac{1}{X}}\qquad(10\cdot3)$$

例：有三种商品，一种是每千克 1.00 元，一种是每千克 0.80 元，一种是每千克 0.50 元，现在各买 1 元，问平均每千克的价格。用上式计算可得：

$$\overline{X}_H=\frac{n}{\sum\frac{1}{X}}=\frac{3}{\frac{1}{1.0}+\frac{1}{0.8}+\frac{1}{0.50}}=0.71(\text{元}/\text{千克})$$