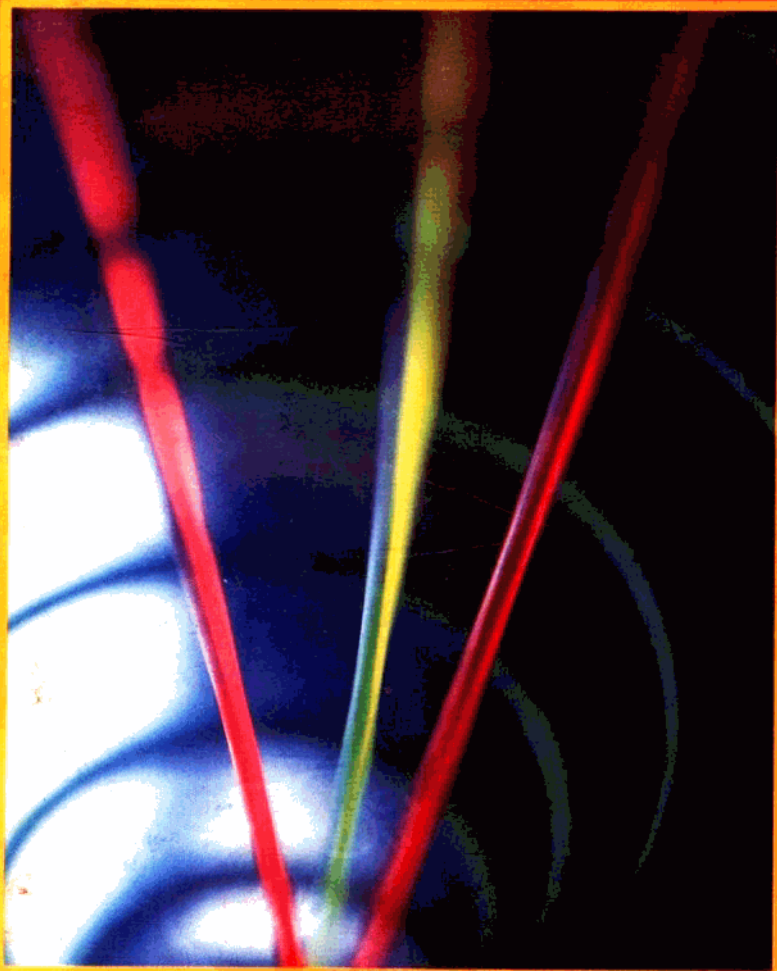


MBA — 工商精英速成丛书

商务统计与预测



目 录

1. 运筹帷幄——商务统计概述…………… (1)
 - 1.1 统计的概念和用途…………… (2)
 - 1.2 统计的基本步骤…………… (4)
 - 1.3 频率分布图…………… (11)
 - 1.4 频率曲线的特征描述…………… (14)

2. 力争上游——度量中心趋势…………… (20)
 - 2.1 算术平均数…………… (22)
 - 2.2 加权平均数…………… (27)
 - 2.3 几何平均数…………… (31)
 - 2.4 中位数和众数…………… (34)

3. 成自险出——概率论及概率分布…………… (38)
 - 3.1 概率的基本定义…………… (40)
 - 3.2 概率的基本运算法则…………… (44)
 - 3.3 全概率公式和逆概率公式…………… (52)
 - 3.4 随机变量的概率分布…………… (56)

4. 管中窥豹——抽样和抽样分析…………… (73)
 - 4.1 随机抽样…………… (75)
 - 4.2 抽样分布…………… (79)
 - 4.3 样本容量与抽样平均误差的关系…………… (91)

5.	审时度势——估计	(93)
5.1	点估计	(95)
5.2	区间估计	(98)
5.3	应用 t 分布做出区间估计	(106)
5.4	确定适当的样本容量	(110)
6.	去伪存真——假设检验	(116)
6.1	假设检验的意义	(116)
6.2	总体平均数的假设检验	(121)
6.3	两个总体平均数差异的假设检验	(125)
7.	高瞻远瞩——回归分析和相关分析	(131)
7.1	一元线性回归方程	(133)
7.2	相关分析	(143)
7.3	总体参数的推断	(146)
7.4	多元回归分析和多元相关分析	(149)
8.	因势利导——时间数列	(157)
8.1	长期趋势分析	(160)
8.2	周期波动分析	(167)
8.3	季节变动分析	(170)
8.4	不规则变化	(178)
8.5	时间数列分析举例	(178)
9.	报告成绩——指数	(188)
9.1	指数的概念	(189)

9.2	综合指数	(192)
9.3	相对数平均法	(200)
10.	决胜千里——决策论	(207)
10.1	最大期望利润或最小期望损失	(210)
10.2	边际分析法	(215)
10.3	决策树分析法	(221)
11.	学以致用——统计学在预测技术中的应用	(230)
11.1	移动平均数	(231)
11.2	分解法	(236)
11.3	线性模型	(236)
附录 1	二项分布表	(241)
附录 2	泊松概率表	(250)
附录 3	标准正态分布表(Z 值表)	(256)
附录 4	t 分布表(t 值表)	(257)
	编后记	(259)

运筹帷幄

——商务统计概述

“运筹帷幄”一词出自《汉书·高帝纪》：“上（指刘邦）曰：夫运筹帷幄之中，决胜千里之外，吾不如子房（张良，字子房）。”刘邦带兵平定天下，多亏张良帮助他出谋划策，制定正确的战略。因此，刘邦说了这段话称赞他。于是，运筹帷幄这个词就意味着指挥员制定作战策略。同样，企业经理也需要制定经营策略，以取得经营的成功。但是，经营策略不是凭空想象的，而是在对客观情况充分了解的基础上做出来的。对客观情况一无所知，凭空决断，靠碰运气去取胜，这在战场上和企业经营中都是不可取的。我国最早的军事家孙子指出：行军打仗必须预先知道战争双方的条件，要做到知己知彼，要算计战争发展的可能结果。作为指挥员，多算则多胜，少算少胜，不算则不胜。这里的“算”就是指信息数据收集和处理过程。从客观世界中收集的信息和数据，必须经过处理，才具有价

值，才能有助于决策者做出适当的决策。在现代经营管理中，统计学是收集和处理数据的一门学科。为经理们运筹帷幄，提供了科学的方法。因此，企业经理必须学习并掌握统计学的概念和数据处理方法，才能根据已知的有限信息，去推断事物发展的可能结果。

1.1 统计的概念和用途

1985年7月1日，中华人民共和国第一次准确的公布：“中国大陆人口总数为10亿7千万。”你知道这个数字是怎么得来的吗？也许你会说，这很简单，从基层到中央，逐级登记累加，就得出这个数字了嘛！事实的确如此。而这个“逐级累加”过程，就是我们这本书所讲的统计的原始含义。

对于不同的人来说，“统计”具有不同的含义和用途。例如，中华人民共和国为了准确地知道全国的实际人口，“统计”就是求出全国的实际人口总数。这对国家的经济发展规划、生产、教育和社会福利等各方面都具有极为重要的意义。对于教师来说，“统计”就是记录全班各个学生的学习成绩和全班的平均成绩，并且以此来评价这个班的学习成绩优劣，评价教师本人的教学效果。对于一个商品销售公司的经理来说，“统计”就是获取每天的商品销售量、库存量、销售额等方面的报告，以便作出相

应的决策。

在客观世界里，存在着两类基本现象，一类是确定性现象，一类是不确定性现象。所谓确定性现象，就是在一定的条件下，现象的结果是必然的，即必然发生或者必然不发生，而且在事先我们可以把握其结果。例如，在一个标准大气压和 100°C 温度的条件下，水一定沸腾。所谓不确定性现象，就是在一定的条件下，现象的结果不只是一个，可能这样出现，也可能那样出现，事先人们是无法肯定的。例如，一个人的死因是多种多样的，可能是病死，也可能是车祸，或者别的什么原因。

统计学就是以不确定性现象为主要研究内容，并把确定性现象作为不确定性现象的特例来处理。

统计学是一门以概率论为基础，研究不确定性现象的数据资料收集、整理、分析和推断的学科。当今，统计学已广泛地应用于社会科学、自然科学、技术科学和管理科学等领域。计算机的发展和普及，为统计学提供了先进、精确、高速的数据处理手段，大大地拓宽了统计学的实用性。在企业经营管理领域，统计学已经成为管理者制定经营策略，预测产品销售量和评价企业经营状况的不可缺少的科学工具。

犹如其它学科一样，统计理论和实践的发展经过了一个相当长的历史过程。在西方现代企业经营中，经理人员已经普遍的应用了统计技术。根据统计技术的不同作用，统计学家通常把统计学分为两

大类，即描述统计学和推断统计学。

描述统计学在于描述某种现象。例如，某工厂厂长组织全体工人进行业务考核，如果他计算某个车间的平均成绩是用以表示这个车间工人的业务水平，而不是借以确定其它几个车间的一般状况，那么，这个厂使用的是描述统计学。描述统计学运用表格、线图去反映所描述的现象，使之更直观，更容易理解。

推断统计学则是依据某种现象，推断其可能出现的结果。例如，这位厂长计算出某个车间的工人的业务考核平均成绩，然后用这个平均成绩来估计全厂其它车间的工人的业务水平，他就要应用推断统计学的方法和技术。推断统计学必须把被调查的具体状况加以一般化，才能得出具有普遍性的结论。这个一般化过程，并不要求一定要有百分之百的真凭实据，但必须说明推断结论的真实性程度。因此，推断统计学必须包括推断结论和结论的有效性判断两个方面。推断统计学在决策中得到充分的应用，是企业经理人员重要的管理手段之一。

1.2 统计的基本步骤

企业经理人员做统计工作，一般要经过以下几个步骤：

1. 提出问题

人们所进行的各种活动，总是具有一定的目的。如，人吃饭是为了补充身体消耗的能量；病人去医院是为了看病；企业经理人员搞统计是为了解决问题。一个企业面临的问题是多种多样的。解决不同的问题，所需要的信息资料也不同。因此，统计工作的第一步，就是确定企业目前要解决的问题。例如，一家香水公司规定，每瓶香水净重 16 克，允许有一个很小的误差范围，为了保护顾客的利益和公司的信誉，使每瓶香水的净重与规定不至相差太多，公司经理必须采用统计的方法观察所产香水的净重，一旦发现香水重量不符合规定，就应该采取相应措施，纠正误差，保证顾客的利益不受损害，同时也为公司赢得信誉。

2. 搜集资料

资料可以来自实地观察或实验，也可以取自日常管理的记录。当今，各个企业已普遍使用计算机来存储大量的资料，一旦需要，计算机能迅速地将人们所需要的资料整理成表格、线图或统计数字等，为人们提供决策依据。但是，当我们使用资料时，必须确保所用资料的可靠性，杜绝“垃圾进，垃圾出，”的情况。

统计工作者，一般是通过从总体中抽取样本的途径来搜集资料，根据样本提供的信息推断总体状况。例如，一个完整的蛋糕是一个总体，你要想知道蛋糕的味道是否可口，完全没有必要吃完整个蛋糕，而只需切下一小块吃了，就能品尝出蛋糕的味

道。香水公司的经理要想知道每瓶香水的净重，把每天生产的香水拿来逐瓶检验是不可能的。因此，他只有从每天生产的全部香水中，抽取一定数量的香水作为样本，检验样本中每瓶香水的重量，然后根据此样本数据，推断出这一天生产的全部香水的平均净重。表 1-1 是某天抽取 30 瓶香水作为样本的净重，这就是经理搜集的原始资料。所谓原始资料，就是未经过统计方法加工的资料。当然，抽取的样本必须能够代表总体。关于抽样方法，我们将在后面专门介绍。

表 1-1 30 瓶香水的净重(单位: 克)

16.2	15.8	15.8	15.8	16.3	15.6
15.7	16.0	16.2	16.1	16.8	16.0
16.4	15.2	15.9	15.9	15.9	16.8
15.4	15.7	16.0	15.9	16.3	16.0
16.4	16.6	15.6	15.6	16.9	16.3

3. 整理数据数列

数据数列就是把搜集的样本数据，按大小（从大到小或从小到大）排列所得到的统计数字表。我们把香水公司抽取的 30 瓶香水的样本数据，从小到大排列，就得到表 1-2 所示的数据数列表。

数据数列与原始数据相比较，有以下优点：

- (1) 便于看出数据中的最大值和最小值，易

表 1-2 30 瓶香水净重的数据数列(单位: 克)

15.2	15.7	15.9	16.0	16.2	16.4
15.4	15.7	15.9	16.0	16.3	16.6
15.6	15.8	15.9	16.0	16.3	16.8
15.6	15.8	15.9	16.1	16.3	16.8
15.6	15.8	16.0	16.2	16.4	16.9

于确定数据的范围。表 1—2 显示了 30 瓶香水的净重值, 范围在 15.2 克到 16.9 克之间。

(2) 便于分组, 有利于进一步的研究。因为当经理人员要编制频率分布以及绘制图表时, 还需要将数据按一定要求分组。数据数列为分组提供了极大的方便。

(3) 便于发现数列中每个重量值出现的次数。从表 1—2 可以看出, 有 9 个重量值的出现都多于一次。

(4) 可以看出两个相邻数值之间的距离。例如表 1—2 中, 15.8 与 15.9 是两个相邻数值, 其距离为 0.1。

数据数列的最大缺点就是太详细, 太繁琐, 人们很难据此总结出一般化特征。比如说, 香水公司经理抽取的样本数不是 30 瓶, 而是 30000 瓶, 那么 30000 个样本数据排列成一个数据数列, 那就太繁琐了, 以致失去了研究的意义。因此, 这就需要略去和压缩一些意义不大的细节信息, 并仍能说明

现象的特征。

4. 编制频率分布

压缩信息的一种有效方法，就是编制频率分布。频率分布把数据分成若干个组，这些组就是描述现象的某个特征的数值群。在全部观察数据中，落入某个组的数据个数，就叫做频率。某个组的观察值个数在全部观察值个数中所占的比重，称为相对频率。相对频率常以分数、小数、或百分数表示。表 1—3 是 30 瓶香水净重的频率分布。

表 1—3 30 瓶香水净重的频率分布(组距: 0.5 克)

组别(克)	频 率	相对频率
(1)	(2)	(3) = (2) ÷ 30
15.1—15.5	2	0.0667
15.6—16.0	16	0.5333
16.1—16.5	8	0.2667
16.6—17.0	4	0.1333
合 计	30	1.0000

根据原始数据编制频率分布，一般分三个步骤：

(1) 确定分组标志，分组数目和组距

对于不同的问题，分组标志也不同。如果我们要考察某地区人口的年龄状况，则岁数（年龄）是分组标志。在前面的例子中，是香水公司经理要考

察所产香水的净重，因此，每瓶香水的净重值——克数，就是分组标志。还另外有颜色、形状、长度等，都是不同的分组标志。

分组数目的多少，取决于资料中数据的多少和数值极差（即最大值与最小值之差）的大小。数据点愈多，或极差愈大，分组的数目就愈多。统计工作者习惯使用的分组数目，一般不少于6个，不大于15个。

$$\text{组距} = \frac{\text{最大值的下一个单位值} - \text{最小值}}{\text{分组数目}} \quad (1-1)$$

香水公司的数据共分为6组，最大值是16.9，下一个单位值是17.0，则：

$$\text{组距} = \frac{17.0 - 15.2}{6} = 0.3(\text{克})$$

由第一步，我们已经确定了香水公司的统计数据的分组标志——香水净重克数；分组数目——6组；组距——0.3克。

(2) 按确定的组数和组距，把各样本数据分组，并计算其频率分布，如表1-4所示。显然，每个样本数据都归入了对应的数据组中。同一个数据不会归入不同的组里。所以分组既无遗漏，又相互独立。

表 1-4 30 瓶香水净重的频率分布(组距: 0.3 克)

组别	频 率	相对频率
15.2—15.4	2	0.0667
15.5—15.7	5	0.1667
15.8—16.0	11	0.3666
16.1—16.3	6	0.2000
16.4—16.6	3	0.1000
16.7—16.9	3	0.1000
合 计	30	1.0000

(3) 绘制频率分布图

用横坐标表示香水的净重，纵坐标表示频率，就可绘制成如图 1-1 所示的频率分布图。

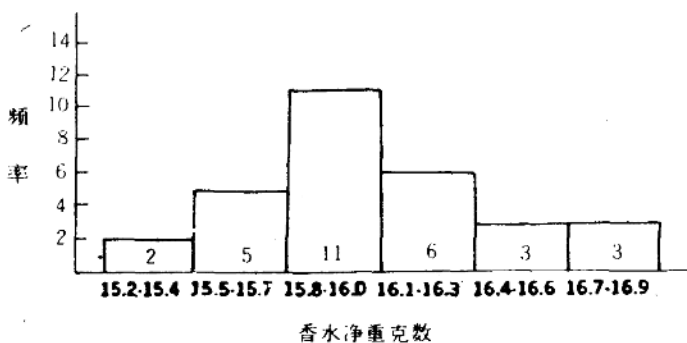


图 1-1 30 瓶香水净重的频率分布

频率分布图直观地反映了样本数据的变动趋

势。如果分组数目愈多，组距愈窄，则频率分布图就愈趋于一条光滑的曲线。但分组数目过多，又会失去分组的意义。如果分组数目愈少，组距愈宽，频率分布图就愈集中。但组距过宽，又会失去很多信息，结果分布图不能说明任何问题，掩盖了数据的变动趋势。

1.3 频率分布图

为了便于直观地认识所描述的现象的特征，人们通常将所获得的被研究对象的数据，绘制成频率分布图。它能突出地反映数据变动趋势。频率分布图有助于人们理解频率分布，根据它可以推测某些数值，并可以在图上检验所做结论的准确性。

下面介绍几种频率分布图及其绘制方法。

1. 直方图

图 1-1 就是直方图。直方图由若干个长方形组成。每个长方形的宽度取决于组距。若采用等距分组，每个长方形的宽度相等。如果是绝对频率直方图，则每个长方形的高度对应着该组数据个数（也就是绝对频率）。如果是相对频率直方图，则长方形的高度对应着该组的相对频率。相对频率直方图与绝对频率直方图的形状完全相同，所不同的是纵坐标表示的意义不同。图 1-2 为香水公司经理绘制的相对频率直方图。

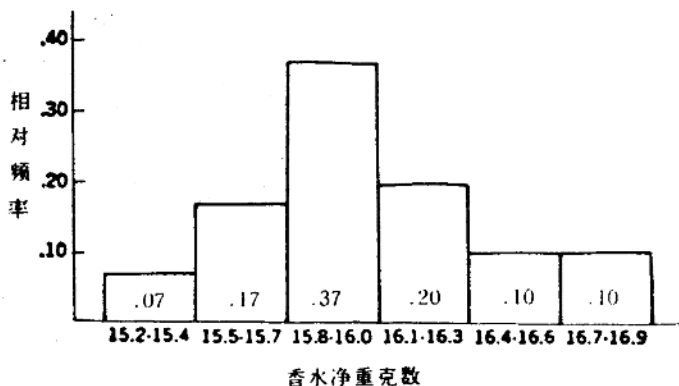


图 1-2 30 瓶香水净重的相对频率直方图

2. 频率多边形

在直角坐标系横轴上，确定每组数据的组中值，再根据组中值向上平行于纵轴描出各组频率对应的点，然后用一条折线连接各点，就绘成了频率多边形。所谓组中值，就是每组数据的中间值。例如，有一组数为 1, 3, 5, 8，则组中值为 4.5，其计算公式如下：

$$\text{组中值} = \frac{\text{数据组的下限} + \text{数据组的上限}}{2} \quad (1-2)$$

例如，香水公司关于香水净重的数据组有 6 个，第一组数据的下限为 15.2 克，上限为 15.4 克，则：

$$\text{组中值} = \frac{15.2 + 15.4}{2} = 15.3$$

表 1-5 列出了 30 瓶香水净重值的分组及各组

的组中值。图 1-3 就是香水公司香水净重的频率多边形。频率多边形也可以变换成直方形，如图 1-3 的虚线所示。

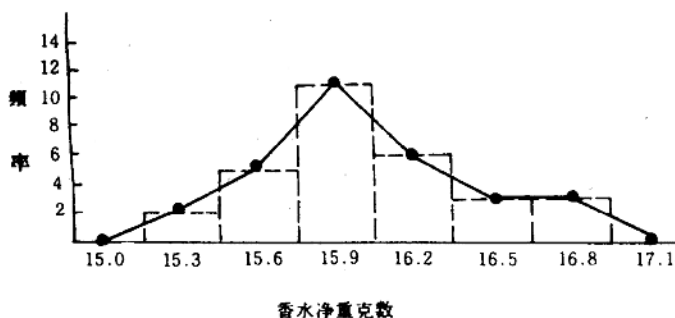


图 1-3 30 瓶香水净重的频率多边形

表 1-5 30 瓶香水净重的分组及组中值

组别	组中值	频率
15.2—15.4	15.3	2
15.5—15.7	15.6	5
15.8—16.0	15.9	11
16.1—16.3	16.2	6
16.4—16.6	16.5	3
16.7—16.9	16.8	3
合计		30

人们比较习惯阅读曲线图。那么怎样才能使直