



普通高等院校经济管理类“十三五”应用型规划教材
【经济管理类专业基础课系列】

第3版

统计学

STATISTICS (3rd Edition)

张兆丰 编著

免费提供
授课用
电子课件



机械工业出版社
China Machine Press



扫描全能王 创建

第4章 统计指标

学习目标

1. 理解各类统计指标的概念和分类方法。
2. 理解各类统计指标的作用和特点,理解其表示的经济意义。
3. 熟练掌握各类指标的计算方法。

统计学是从数量方面认识社会经济现象运行的规律性,而其中一种基本方法就是利用统计指标对现象特征进行描述,这种方法称为综合指标法。统计指标包括总量指标、相对指标、平均指标和变异指标,这些指标从不同的侧面反映了总体的数量特征和数量关系,是我们认识社会经济现象的重要途径。

4.1 总量指标

4.1.1 总量指标的含义

总量指标是反映社会经济现象在一定时间、空间条件下总规模、总水平或工作总量的综合指标,一般用绝对数表示,又称绝对数指标或绝对指标。例如,2018年2月28日国家统计局发布了2017年经济数据:经初步核算,全年国内生产总值827 122亿元。其中,第一产业增加值65 468亿元;第二产业增加值334 623亿元;第三产业增加值427 032亿元。2017年年末全国内地总人口(包括31个省、自治区、直辖市的人数和中国人民解放军现役军人,不包括香港、澳门特别行政区和台湾地区以及海外华侨人数)139 008万人,比2016年年末增加737万人,全年出生人口1 723万人。全国人户分离的人口2.91亿人,其中流动人口2.44亿人。2017年年末全国就业人员77 640万人,其中城镇就业人员42 462万人。全年城镇新增就业1 351万人,比2016年增加37万人。全国农民工总量28 652万人,其中外出农民工17 185万人,本地农民工11 467万人。2017年年末国家外汇储备31 399亿美元,比2016年年末增加1 294亿美元。全年粮食产量61 791万吨。其中夏粮产量14 031万吨;



早稻产量 3 174 万吨；秋粮产量 44 585 万吨。2017 年全社会固定资产投资 641 238 亿元，其中固定资产投资（不含农户）631 684 亿元。这些指标都是总量指标，都是利用绝对数说明我国 2013 年国民经济发展的总体规模、总体水平和全国人民的生活水平的指标。

总量指标由六个要素构成：名称、时间、空间、数值、计量单位和计算方法。例如，国家统计局发布的人口指标，名称是总人口；时间是 2017 年年末；空间是 31 个省、自治区、直辖市的人数和中国人民解放军现役军人，不包括香港、澳门特别行政区和台湾地区以及海外华侨人数；数据是 139 008；计量单位是万人；计算方法是登记汇总。

总量指标也可以表现为社会经济现象总体在一定时空条件下数量增减变化的绝对数。例如，2017 年年末全国大陆人口数总比上年年末增加 1 351 万人，也属于总量指标。

4.1.2 总量指标的计量单位

1. 实物量单位

实物量单位是根据现象的自然属性和特点而采用实物单位计量。实物量单位有自然单位，如人口以人为单位，牲畜以头为单位；有度量衡单位，如粮食以吨为单位，棉布以米为单位；有标准实物量单位，如将含热量不同的煤折合为 7 000 大卡/千克的标准煤；有复合单位，如货物周转量以吨公里为单位等。用实物单位计算的总量指标称为实物指标。

2. 价值量单位

价值量单位以货币为计量单位来度量总体的数量，如 GDP、销售额、利税总额等。用货币单位计算的总量指标称为价值指标，可以综合反映具有不同使用价值的经济现象的总规模、总水平，具有广泛的综合性和概括能力。

3. 劳动量单位

劳动量单位是以劳动过程中消耗的劳动时间为计量单位，如工时、工日等，为成本核算和计算劳动生产率提供依据。用劳动量单位计算的总量指标称为工作量指标或劳动量指标。

4.1.3 总量指标的作用

1. 总量指标可以反映一个总体的基本情况

人们要想正确认识一个总体的基本情况，首先要掌握总体在一定时间、条件下的规模或水平，如人口数、劳动力数量、土地面积等。所以，总量指标是认识客观现象的起点，可以反映一个主题的基本情况，是正确认识国情、国力的起点。

2. 总量指标是制定政策和编制计划、分析各种指标的基础指标

总量指标是国家实现宏观调控和企业进行经营管理的基本指标。在市场经济条件下，



要使国民经济协调发展,就需要对经济运行实行调控;要使企业生产经营活动正常进行,需要实行科学的管理,这就需要我们掌握宏观经济和微观经济运行的环境、条件、投入、产出等各方面的数量状况,研究各方面的数量关系。总量指标可以反映这些现象的数量关系,为经济管理提供依据。

3. 总量指标是计算相对指标、平均指标等各种分析指标的基础

统计综合指标中的相对指标、平均指标的计算都是以总量指标为基础计算的,他们最基本的表现形式都是两个总量指标对比的结果,是总量指标的派生指标。总量指标计算是否科学直接影响到其他指标计算的正确性。

4.1.4 总量指标的种类

总量指标包含时间要素,而经济现象的特性不同,其时间属性也不同。根据总量指标的时间属性不同,可以将其划分为时期指标和时点指标。

时期指标的时间属性是时期,即一段时间,是反映社会经济现象在某一时期内的发展变化过程的总量指标,如商品销售额、总产值、基本建设投资额等。

时点指标的时间属性是时点,即某一瞬间,是反映社会经济现象在某一时点的状况的总量指标,如人口数、存款余额、商品库存量等。

时期指标和时点指标有以下区别。

- 1) 时期指标是经常登记取得的,而时点指标是间断登记取得的。
- 2) 时期指标数值可以直接相加,而时点指标直接相加则无实际意义。
- 3) 时期指标数值大小与时期长短有直接关系,而时点指标则与时间间隔无直接关系。

时期指标和时点指标的统计处理与应用不同,在运用时期指标和时点指标时,注意若从不同的角度考虑同一指标,则总量指标的性质也不同,如年末人口数和年初人口数是时点指标,但年末人口数—年初人口数=人口净增数,人口净增数为时期指标。

4.1.5 应用总量指标注意的问题

总量指标在计算方法上比较简单,但在计算内容上却是相当复杂,因此总量指标的计算并不是一个单纯技术性的加总问题,而必须明确规定总量指标所表示的各种社会经济现象的概念、构成内容和计算范围,确定计算方法,然后才能进行计算汇总,以取得正确反映社会经济现象的总量资料。例如,要正确计算工资总额,必须先明确工资的实质和构成;要计算国民经济各部门职工人数,不仅要明确职工的概念和范围,而且要从理论上先确定国民经济部门的分类,才能得出按部门分类的职工人数。另外,总量指标计算也应注意历史条件变化对指标内容和范围的影响,使指标具有可比性,以利于动态研究。

4.2 相对指标

要全面深入分析一种社会经济现象,仅仅利用总量指标是远远不够的。如果要对事物



做深入的了解,发现现象之间的联系,就需要对总体的组成和其各部分之间的数量关系进行分析、比较,这就必须计算相对指标。

4.2.1 相对指标的概念

相对指标是两个有联系的指标进行对比的比值来反映社会经济现象数量特征和数量关系的综合指标。相对指标将对比的绝对差异抽象化,也称作相对数,其数值有两种表现形式:无名数和复名数。无名数是一种抽象化的数值,多以系数、倍数、番数、成数、百分数或千分数表示。复名数主要用来表示强度的相对指标,以表明事物的密度、强度和普遍程度等。例如,人均粮食产量用“千克/人”表示,人口密度用“人/平方千米”表示等。

这里对经济分析中经常用到的“成数”和“百分点”的概念做一点说明。成数以十作为比较基准的比率,一成是十分之一。例如,春节期间商品销售额增长一成,说明春节期间商品销售额增长十分之一,即10%。百分点是指不同时期以百分数形式表示的相对指标的变动幅度。例如,2013年我国第三产业比重明显提高,达到46.1%,比上年提高1.5个百分点,比第二产业比重高2.2个百分点,这是第三产业比重首次超过第二产业。在很多情况下我们还要考虑一个百分点所对应的绝对数。

4.2.2 相对指标的作用

相对指标在认识社会经济现象中有着非常重要的作用。

1) 相对指标通过数量之间的对比,可以表明事物相关程度、发展程度,它可以弥补总量指标的不足,使人们清楚了解现象的相对水平和普遍程度。例如,某企业去年实现利润500万元,今年实现550万元,则今年利润增长了10%,这是总量指标不能说明的。

2) 把现象的绝对差异抽象化,使原来无法直接对比的指标变为可比。不同的企业由于生产规模条件不同,直接用总产值、利润比较评价意义不大,但如果采用一些相对指标,如资金利润率、资金产值率等进行比较,便可对企业生产经营成果做出合理评价。

3) 说明总体内在的结构特征,为深入分析事物的性质提供依据。例如,计算一个地区不同经济类型的结构,可以说明该地区经济的性质;计算一个地区的第一、第二、第三产业的比重,可以说明该地区社会经济现代化程度等。

4.2.3 相对指标的种类及其计算方法

相对指标由于研究的目的不同,比较的基数不同,可以分为结构相对指标、比例相对指标、比较相对指标、强度相对指标、动态相对指标和计划完成相对指标。

1. 结构相对指标

结构相对指标也称比重,是利用分组法将总体分为不同性质的各个部分,再将各个部



分指标数值与总体对应的指标数值进行对比,说明总体内部组成情况的相对指标,一般用百分数(%)表示。它经常用来分析现象的内部构成情况,把不同时间的结构相对指标进行对比分析,可以发现现象的变化过程和规律;总体各组的结构相对指标可以发现该组在总体的地位和作用,对于计算平均指标有特殊意义。

$$\text{结构相对指标} = \frac{\text{总体某部分数值}}{\text{总体全部数值}} \times 100\% \quad (4-1)$$

【例 4-1】 根据表 4-1 中的数据计算相应的结构相对指标。

表 4-1 2017 年我国 GDP 总值构成情况

国内生产总值	绝对额(亿元)	比重(%)
第一产业	65 467.6	7.92
第二产业	334 622.6	40.46
第三产业	427 032	51.63
合计	827 122.2	100.00 ^①

①由于四舍五入的原因,合计不一定为 100%。

资料来源:中国统计年鉴 2018。

解:根据式(4-1):

第一产业占 GDP 比重为: $(65\,467.6 \div 827\,122.2) \times 100\% \approx 7.92\%$

第二产业占 GDP 比重为: $(334\,622.6 \div 827\,122.2) \times 100\% \approx 40.46\%$

第三产业占 GDP 比重为: $(427\,032 \div 827\,122.2) \times 100\% \approx 51.63\%$

从表 4-1 中的资料可以看出产业构成特点:2017 年我国第三产业在 GDP 中所占比重最大,超过 50%,比第二产业比重高 11.17 个百分点;而第一产业所占比重最低,占比不到 8%。这些特点说明我国经过改革开放,正在逐步向发达的经济形态迈进。

结构相对指标在社会经济统计中结构相对数应用广泛,可以说明在一定的时间、空间条件下,总体结构的特征;通过不同时期结构相对指标的变化,可以反映事物性质的发展趋势,分析经济结构的演变规律;根据各构成部分所占比重大小,可以反映所研究现象总体的质量以及人、财、物的利用情况,有助于分清主次,确定工作重点。

2. 比例相对指标

比例相对指标是利用总体内部各组成部分之间的联系对不同组成部分的指标数值进行对比,可以表明总体内部的比例关系。

$$\text{比例相对指标} = \frac{\text{总体某部分指标数值}}{\text{总体另一部分指标数值}} \quad (4-2)$$

比例相对指标一般表现将总体分成两部分后各部分之间的比例关系。例如,我国第六次人口普查结果显示,在中国内地 31 个省、自治区、直辖市和现役军人的人口中,男性人口为 686 852 572 人,占 51.27%;女性人口为 652 872 280 人,占 48.73%。总人口性别比(以女性为 100,男性对女性的比例)由 2000 年第五次全国人口普查的 106.74 下降为 105.20,即男:女=105.20:100。



有时,比例相对指标也可以表示一个总体分成多个部分后各部分之间的比例关系,这时采用连比形式。

【例 4-2】 根据表 4-1 中的数据,计算第一产业、第二产业、第三产业 GDP 的比例相对指标。

解:根据式(4-2),以第一产业为单位 1 得:

$$\text{第一产业:第二产业:第三产业} = 7.92 : 40.46 : 51.63 = 1 : 5.1 : 6.5。$$

根据统计资料,计算各种比例相对数,反映有关事物之间的实际比例关系,有助于我们认识客观事物是否符合按比例协调发展的要求,参照有关标准,可以判断比例关系是否合理。在宏观经济管理中,这对于研究分析整个国民经济和社会发展是否协调均衡具有重要的意义,可以促使经济稳步协调发展。

3. 比较相对指标

客观事物的发展往往是不平衡的,比较相对指标是同一时间不同国家、不同地区、不同单位的某项指标对比的结果。

$$\text{比较相对指标} = \frac{\text{甲总体某类指标数值}}{\text{乙总体同类指标数值}} \quad (4-3)$$

比较相对指标一般用倍数表示,有时也可用系数表示,计算比较相对指标的指标数值可以是总量指标,也可以是相对指标或平均指标。

【例 4-3】 国际咖啡组织的数据显示,美国是全球咖啡消费最大的市场,年消费额约为 3 万亿元人民币。2017 年,美国咖啡消费量为 24 990 千袋,共 1 499 400 公斤;而巴西咖啡工业协会(Abic)公布的数据表明,2017 年巴西已成为仅次于美国的全球第二大咖啡消费国。Abic 数据显示,巴西国内咖啡消费量为 2 150 万袋,相当于 107 万吨咖啡,仅次于美国。求 2017 年美国与巴西咖啡消费量比较相对指标。

解:由于两个组织所使用的单位不同,所以先要统一单位。美国咖啡消费量为 24 990 千袋,换算为 2 499 万袋。

由式(4-3)得:

$$\text{相对指标} = \frac{\text{美国咖啡消费量}}{\text{巴西咖啡消费量}} = \frac{2\,499}{2\,150} = 1.16$$

或者

$$\text{相对指标} = \frac{\text{巴西咖啡消费量}}{\text{美国咖啡消费量}} = \frac{2\,150}{2\,499} = 0.86$$

计算结果说明,2017 年美国咖啡消费量约为巴西的 1.16 倍,或者巴西咖啡消费量约为美国的 86%。

在经济管理工作中,广泛应用比较相对数。例如,用各种指标在企业之间、车间或班



组之间进行对比,把各项技术经济指标与国家规定的标准条件对比,与同类企业的先进水平或世界先进水平对比,借以找出差距,挖掘潜力,定出措施,为提高企业的经营管理水平提供依据。

计算比较相对数应注意对比指标的可比性,包括指标的计算口径、所属的时间、计量单位等。此外,比较基数的选择要根据资料的特点及研究目的而定。如上例中以巴西咖啡消费量作为比较基数,计算结果说明美国咖啡消费量是巴西的1.16倍;如以美国咖啡消费量作为比较基数,则表明巴西咖啡消费量是美国的86%。这两种计算方法的角度不同,但都能说明问题,具体以哪个指标作为比较的基础,应根据研究目的以及哪种方法能更确切地说明问题的实质而定。

4. 强度相对指标

强度相对指标是两个性质不同而有联系的总量指标对比的结果,是反映现象的强度、密度、普及程度的指标。

$$\text{强度相对指标} = \frac{\text{某一总量指标数值}}{\text{另一个有联系而性质不同的总量指标数值}} \quad (4-4)$$

强度相对指标有些是以复名数表示的,也有些是采用无名数。由于强度相对数是两个性质不同但有联系的总量指标数值之比,所以在多数情况下,是由分子与分母原有单位组成的复合单位表示的,如人口密度用人/平方千米,人均钢产量用吨/人等。但有少数的强度相对指标因其分子与分母的计量单位相同,可以用千分数或百分数表示其指标数值,如人口自然增长率使用千分数。

【例 4-4】随着我国经济的发展,医疗卫生事业也保持平衡较快发展。描述医疗卫生机构密度的一个指标是“每千人口医疗卫生机构床位”,这个指标的计算公式为:

$$\text{每千人口医疗卫生机构床位} = \frac{\text{年末医疗卫生机构床位数}}{\text{年末人口数}}$$

根据国家统计局发布的《统计年鉴 2018》,近几年我国“每千人口医疗卫生机构床位”的数据如表 4-2 所示。

表 4-2 每千人口医疗卫生机构床位

年份	每千人口医疗卫生机构床位		
	合计	城市	农村
2015	5.11	8.27	3.71
2016	5.37	8.41	3.91
2017	5.72	8.75	4.19

如表 4-2 所示,每千人口医疗卫生机构床位逐年提高,说明我国医疗卫生事业在不断发展,但是城市和农村的发展不平衡。

利用强度相对数来说明社会经济现象的强弱程度时,广泛采用人均产量指标来反映一



个国家的经济实力。例如,按全国人口数计算的人均钢产量、人均粮食产量等,这种强度相对指标的数值越大,表示一个国家的经济发展程度越高,经济实力越强。

有些强度相对指标的分子和分母可以互换,既可采用正指标,也可采用逆指标。所谓正指标是指指标数值越大,现象强度越高;反之就是逆指标。例如:

$$\text{银行网点密度(正指标)} = \frac{\text{银行机构数(个)}}{\text{地区人口数(千人)}}$$

$$\text{银行网点密度(逆指标)} = \frac{\text{地区人口数(千人)}}{\text{银行机构数(个)}}$$

特别需要说明的是,强度相对指标在很多情况下带有“平均”的含义,但是,强度相对指标不是平均指标,两者的差别我们会在平均指标这部分内容里加以分析。

5. 动态相对指标

动态相对指标就是将同一现象在不同时期的两个数值进行动态对比而得出的相对数,说明现象在不同时间上的发展方向和变化程度,一般用百分数或倍数表示,也称为发展速度。其计算公式如下:

$$\text{动态相对指标} = \frac{\text{报告期水平}}{\text{基期水平}} \times 100\% \quad (4-5)$$

通常,我们将需要报告或研究的时期称为报告期,而将作为比较基础的时期称为基期;报告期和基期对应的指标数值称为报告期水平和基期水平。

基期的选择具有主观性。如果基期固定不变,计算得到的是定基动态相对指标;如果以报告期的前一期为基期,计算得到的是环比动态相对指标。

【例 4-5】 表 4-3 是 2010~2017 年我国货物周转量的数据,以报告期的前一期为基期,计算环比发展速度,结果如表 4-3 所示。

表 4-3 我国货物周转量

年份	2010	2011	2012	2013	2014	2015	2016	2017
货物周转量(亿吨千米)	141 837	159 324	173 804	168 014	181 668	178 356	186 629	197 373
环比发展速度(%)	—	112.33	109.09	96.67	108.13	98.18	104.64	105.76

资料来源:中国统计年鉴 2018。

由计算结果可以看出,2013 年和 2015 年我国货物周转量略有下降,其他年份均在增长。

在动态相对指标中,还有一种特殊的比较方法——同比。同比是当月(或)季的指标与去年同期相比的简称,目的是消除周期性的影响。

动态相对指标在统计分析中应用很广,将在第 6 章中详加论述。

6. 计划完成程度相对指标

计划完成程度相对指标是社会经济现象在某时期内实际完成数值与计划任务数值对比



的结果,反映完成任务的程度,一般用百分数来表示。

基本计算公式为:

$$\text{计划完成程度相对指标} = \frac{\text{实际完成数}}{\text{计划任务数}} \times 100\% \quad (4-6)$$

由于掌握的资料不同,在计算时需要対式(4-6)进行变换,下面以例题的形式加以说明。

【例 4-6】 某企业计划当年产值达 100 万元,实际执行结果产值达 115 万元,求计划完成程度。

解:由式(4-6)得:

$$\text{计划完成程度} = \text{实际完成数} \div \text{计划任务数} \times 100\% = 115 \div 100 \times 100\% = 115\%$$

由于 $115\% - 100\% = 15\%$,说明该企业超额 15% 完成计划。

【例 4-7】 某企业计划劳动生产率当年比上年增长 3%,实际增长 5%,求计划完成程度。

解:由式(4-6)得:

$$\begin{aligned} \text{计划完成程度} &= \text{实际完成数} \div \text{计划任务数} \times 100\% \\ &= (1 + 5\%) \div (1 + 3\%) \times 100\% = 101.94\% \end{aligned}$$

由于 $101.94\% - 100\% = 1.94\%$,说明该企业超额 1.94% 完成计划。

【例 4-8】 某企业计划废气排放当年比上年下降 3%,实际下降 5%,求计划完成程度。

解:由式(4-6)得:

$$\begin{aligned} \text{计划完成程度} &= \text{实际完成数} \div \text{计划任务数} \times 100\% \\ &= (1 - 5\%) \div (1 - 3\%) \times 100\% = 97.94\% \end{aligned}$$

由于 $97.94\% - 100\% = -2.06\%$,说明该企业超额 2.06% 完成计划。

4.2.4 计算和运用相对指标的原则

上述六种相对指标从不同的角度出发,运用不同的对比方法,对同类指标数值进行静态的或动态的比较,对总体各部分之间的关系进行数量分析,对两个不同总体之间的联系程度和比例做比较,是统计中常用的基本数量分析方法之一。要使相对指标在统计分析中起到应有的作用,在计算和应用相对指标时应该遵循以下的原则。

1. 可比性原则

相对指标是两个有关的指标数值之比,对比结果的正确性取决于两个指标数值的可比性。如果违反可比性这一基本原则计算相对指标,就会失去实际意义,导致不正确的结论。对比指标的可比性,是指对比的指标在含义、内容、范围、时间、空间和计算方法等口径方面协调一致,相互适应。如果各个时期的统计数值因行政区划、组织机构、隶属关



系的变更,或统计制度方法的改变不能直接对比,就应以报告期的口径为准,调整基期的数值。

2. 定性分析与定量分析相结合的原则

计算对比指标数值的方法是简便易行的,但要正确地计算和运用相对数,还要注重定性分析与定量分析相结合的原则。因为事物之间的对比分析,必须是同类型的指标,只有通过统计分组,才能确定被研究现象的同质总体,便于同类现象之间的对比分析。这说明要在确定事物性质的基础上,再进行数量上的比较或分析,而统计分组在一定意义上也是一种统计的定性分类或分析。即使是对同一种相对指标在不同地区或不同时间进行比较,也必须先对现象的性质进行分析,判断是否具有可比性。同时,通过定性分析,可以确定两个指标数值的对比是否合理。例如,将不识字人口数与全部人口数对比来计算文盲率,显然是不合理的,因为其中包括未达学龄的人数和不到接受初中文化教育年龄的人数在内,不能如实反映文盲人数在相应的人口数中所占的比重,所以通常计算文盲率的公式为:

$$\text{文盲率} = \frac{\text{15岁以上不识字人口数}}{\text{15岁以上全部人口数}} \times 100\%$$

3. 相对指标和总量指标结合运用的原则

绝大多数的相对指标都是两个有关的总量指标数值之比,用抽象化的比值来表明事物之间对比关系的程度,而不能反映事物在绝对量方面的差别。因此在一般情况下,相对指标离开了据以形成对比关系的总量指标,就不能深入地说明问题。例如,两家企业,当年的利润分别增长5%和8%,但是并不能以此说明增长8%的企业优于另外一家企业,因为这两家企业的规模可能有很大的差异,每增长一个百分点所对应的绝对额可能是不一样的。

4. 各种相对指标综合应用的原则

各种相对指标的具体作用不同,都是从不同的侧面来说明所研究的问题。为了全面而深入地说明现象及其发展过程的规律性,应该根据统计研究的目的,综合应用各种相对指标。例如,为了研究工业生产情况,既要利用生产计划的完成情况指标,又要计算生产发展的动态相对数和强度相对数。由此可见,综合运用结构相对数、比较相对数、动态相对数等多种相对指标,有助于我们剖析事物变动中的相互关系及其后果,更好地阐明现象之间的发展变化情况。

4.3 平均指标

平均指标是反映现象一般水平的指标,也是最能反映现象特征的指标,大量的统计规



律都是以平均指标的形式表现出来的,所以研究平均指标有着非常重要的意义。

4.3.1 平均指标的概念与作用

1. 平均指标的概念

平均指标是反映现象在一定时间、空间条件下所达到的一般水平的综合指标,也称为统计平均数,简称为平均数。

由于总体中的个体在某一变量的取值上存在差异,很多时候我们需要找出一个将这些差异抽象化,代表这个变量取值一般水平的指标,由于总体中的个体具有同质性,所以可以用代表一般水平的指标来反映总体的数量特征,这个指标就是平均指标。它是数据的代表值或中心值,描述了数据分布的集中趋势。取得平均指标的方法通常有两种:一是从总体各单位变量值中抽象出具有—般水平的量,这个量不是个体的具体变量值,而是反映个体的一般水平,这种平均数称为数值平均数。数值平均数有算术平均数、调和平均数、几何平均数等形式。二是先将个体的变量值按—定顺序排列,然后取某—位置的变量值来反映个体的一般水平,把这个特殊位置上的数值看作平均数,称为位置平均数。位置平均数有众数、中位数、四分位数等形式。

2. 平均指标的作用

1) 平均指标可以消除因总体范围不同而带来的总体数量差异,从而使不同的总体具有可比性。例如,由于播种面积不同,不同地区小麦总产量不便直接对比,若计算小麦平均亩产量,则可比较判断不同地区小麦生产水平的高低。

2) 同一总体在不同时间上的平均指标可以反映现象总体的发展变化趋势。例如,研究对比不同时期的职工平均工资,就可正确反映职工工资水平的变化趋势和规律。

3) 利用平均指标可以分析现象之间的依存关系。例如,将学生按成绩分组,计算各组学生的平均成绩与平均学习时间,就可以观察学生的成绩与学习时间之间的依存关系。

4) 平均指标是进行统计推断的一个主要指标。

4.3.2 算术平均数

算术平均数也称均值,是平均指标中最重要的一种,它是在所有平均指标中应用最广泛的平均数。

算术平均数的基本公式为:

$$\text{算术平均数} = \frac{\text{所有数据之和}}{\text{所有个体数量}} \quad (4-7)$$

例如,职工平均工资等于职工的工资总额与职工总人数之比。

算术平均数的分子与分母都是总量指标,分别称为“标志(变量)总量”和“单位(个体)总量”,它们具有“一一对应”的关系,分子依赖于分母而存在。例如,工资(数



据)要依赖于职工(个体)而存在,只有这样计算平均指标才能表明现象的一般水平。

正是在这点上,平均指标与强度相对指标表现出性质上的差异。强度相对指标是两个有联系的不同总体的总量指标对比,这两个总量指标没有依赖关系,而只是在经济内容上存在客观联系。以此标准来衡量,职工平均工资、人均粮食消费量等是平均指标;而人均收入、人均粮食产量是强度相对指标。

平均数分为总体平均数和样本平均数。总体平均数是根据总体各个单位的标志值或标志特征计算的,反映总体一般水平的指标,用字母 μ 表示;样本平均数是根据样本各个标志值或标志特征计算的,反映样本一般水平的指标,用符号 \bar{X} 表示。一般情况下我们只能计算样本平均数。

在实际工作中,就手工计算而言,由于所掌握的统计资料的不同,利用上述公式进行计算时,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

简单算术平均数是根据未经分组整理的原始数据计算得到,设变量值分别为 X_1, X_1, \dots, X_n ,则简单算术平均数的计算公式如下:

$$\bar{X} = \frac{X_1 + X_2 + \dots + X_n}{n} = \frac{\sum_{i=1}^n X_i}{n} \quad (4-8)$$

式中 \bar{X} ——算术平均数;

X_i ——变量值;

n ——变量值的个数。

【例 4-9】 某企业抽取 20 名新入职员工的月工资数据如下(单位:元)。

3 260	3 360	3 300	3 300	3 360	3 400	3 360	3 500	3 260	3 300
3 360	3 360	3 300	3 400	3 360	3 360	3 260	3 300	3 360	3 400

计算新入职员工的人均月工资。

解:根据式(4-8)新入职员工的人均月工资为:

$$\bar{X} = \frac{3\,260 + 3\,360 + \dots + 3\,360 + 3\,400}{20} = 3\,343(\text{元/人})$$

2. 加权算术平均数

(1) 由单项数列计算加权算术平均数 根据分组整理的数据计算算术平均数,要以各组变量值出现的次数或频数为权数计算加权的算术平均数。设原始数据被分成 m 组,各组的变量值为 X_1, X_2, \dots, X_m ,各组变量值的次数或频数分别为 f_1, f_2, \dots, f_m ,则加权的算术平均数为:



$$\bar{X} = \frac{X_1 f_1 + X_2 f_2 + \cdots + X_m f_m}{f_1 + f_2 + \cdots + f_m} = \frac{\sum_{i=1}^m X_i f_i}{\sum_{i=1}^m f_i} \quad (4-9)$$

【例 4-10】 将例 4-9 中的数据分组，得到表 4-4，试计算新入职员工平均月工资。

表 4-4 新入职员工月工资统计表

月工资 (元) X	人数 (人) f	各组工资总额 (元) Xf
3 260	3	9 780
3 300	5	16 500
3 360	8	26 880
3 400	3	10 200
3 500	1	3 500
合计	20	66 860

$$\text{解: } \bar{X} = \frac{\sum Xf}{\sum f} = \frac{9\,780 + 16\,500 + 26\,880 + 10\,200 + 3\,500}{3 + 5 + 8 + 3 + 1} = \frac{66\,860}{20} = 3\,343(\text{元})$$

现在我们分析一下这个计算结果。平均数 3 343 的位置离分组值 3 360 最近，而 3 360 这组的次数是最多的，所以平均数出现在次数最多的变量值附近。而且平均数 3 343 的位置偏向分组值 3 300，这是因为 3 300 组的次数大于 3 400 组的次数。由此可以看出，次数 f 有权衡轻重的作用，故称其为权重，所对应计算得到的平均数也称为加权算术平均数。

将例 4-10 的计算结果用点图表示如图 4-1 所示。

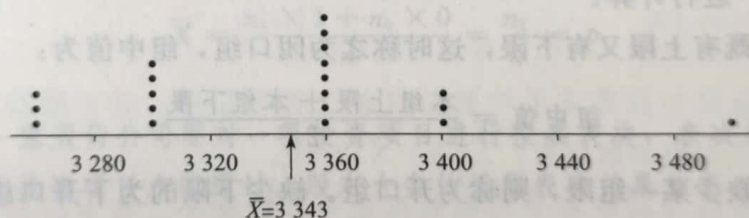


图 4-1 均值的位置示意图

由图 4-1 可以得到一个直观的结论：均值是分布的重心，也就是说均值一定出现在使分布左右平衡的位置上。如果分布发生了改变，均值的位置也将随之改变。

加权算术平均数的另外一种表现形式为：

$$\bar{X} = \frac{\sum_{i=1}^m X_i f_i}{\sum_{i=1}^m f_i} = \sum_{i=1}^m X_i \cdot \frac{f_i}{\sum_{i=1}^m f_i} \quad (4-10)$$

此时的权重是频率。由式 (4-10) 可以看出，加权算术平均数受到两个因素的影响：一个因素是变量值 X ，另外一个因素是频率 $f/\sum f$ 。频率越大，相应的变量值计入平均数的份额也越大，对平均数的影响就越大；反之，频率越小，相应的变量值计入平均数的份额也越小，对平均数的影响就越小。

【例 4-11】 如果表 4-4 中的统计数据是频率的形式（见表 4-5），计算月平均工资。

表 4-5 新入职员工月工资频率统计表

月工资 (元) X	频率 ($f/\sum f$)	月工资比重 ($X \cdot f/\sum f$)
3 260	0.15	489
3 300	0.25	825
3 360	0.40	1 344
3 400	0.15	510
3 500	0.05	175
合计	1.00	3 343

解：由式 (4-10) 得：

$$\bar{X} = \sum_{i=1}^m X_i \cdot \frac{f_i}{\sum_{i=1}^m f_i} = 3\,260 \times 0.15 + 3\,300 \times 0.25 + 3\,360 \times 0.40 +$$

$$3\,400 \times 0.15 + 3\,500 \times 0.05 = 3\,343(\text{元})$$

这个计算结果与例 4-10 中的计算结果完全相同。

如果我们将频率看作对应的变量值出现的可能性 (概率)，则可以将式 (4-10) 计算得到的平均数看作数学期望。

(2) 由组距数列计算加权算术平均数 由于在组距数列中不能具体确定每一个数据值，所以要用各组的组中值代替每组的数据值。其方法是先确定各组的组中值，再用式 (4-9) 或 (4-10) 进行计算。

如果分组资料既有上限又有下限，这时称之为闭口组，组中值为：

$$\text{组中值} = \frac{\text{本组上限} + \text{本组下限}}{2} \quad (4-11)$$

如果分组资料缺少某一组限，则称为开口组。缺少下限的为下开口组，缺少上限的为上开口组。在等距分组的条件下，组中值为：

$$\text{下开口组组中值} = \text{本组上限} - \frac{\text{相邻组组距}}{2} \quad (4-12)$$

$$\text{上开口组组中值} = \text{本组下限} + \frac{\text{相邻组组距}}{2} \quad (4-13)$$

【例 4-12】某企业 100 名职工的工资如表 4-6 所示，试计算平均工资。

表 4-6 某单位职工的月工资情况表

按月工资分组 (元)	人数 (人) f	组中值 X	各组工资总额 Xf (元)
4 000 以下	10	3 750	37 500
4 000~4 500	27	4 250	114 750
4 500~5 000	45	4 750	213 750
5 000~5 500	15	5 250	78 750
5 500 以上	3	5 750	17 250
合计	100	—	462 000

解：根据式 (4-10) 得：



$$\bar{X} = \frac{\sum Xf}{\sum f} = \frac{37\,500 + 114\,750 + 213\,750 + 78\,750 + 17\,250}{100} = 4\,620(\text{元})$$

需要指出的是,根据公式计算平均数时,是用各组的组中值代表各组实际数据的。使用组中值时,假定各组数据在各组中是均匀分布,但实际情况可能与这一假定会有一定的偏差,使得利用分组资料计算的平均数与实际的平均值会产生误差,所以根据公式计算的平均数是实际平均值的近似值。

从计算的结果来看,加权算术平均数 4 620 的位置最接近权重最大的组中值 4 750,并且略小于 4 750。

3. 是非变量的平均数

是非变量是描述现象具有某种属性或不具有某种属性的变量。例如,在描述性别时分为男、女;描述产品质量分为合格、不合格等。是非变量的表现形式是比率,即具有某种属性的个体占全部个体的比重。在总体中, N 表示全部个体数, N_1 表示总体中具有某一属性的个体数, N_0 表示总体不具有这种属性的个体数,则总体比率表示为 $\pi = N_1/N$ 。对于样本而言,样本比率为 $p = N_1/N$ 。

一般情况下,我们将具有某一属性的特征变量值记为 1,将不具有这一属性的特征变量值记为 0。

将是非变量赋值后,可以计算其平均数,样本是非变量平均数的计算公式为:

$$\bar{X} = \frac{n_1 \times 1 + n_0 \times 0}{n} = \frac{n_1}{n} = p \quad (4-14)$$

【例 4-13】 一家股份公司要对一项投资项目进行投票表决,表决规则为需超过 60% 的股东的同意该项投资项目才能通过。股东大会的投票表决结果如表 4-7 所示,判断该项投资项目能否通过?(Y=同意, N=不同意)

表 4-7 股东大会投票表决结果

股东	持有股票数量(万股)	投票结果
A	225	Y
B	312	Y
C	178	Y
D	264	N
E	126	N

解:赋值如下:Y=1, N=0。

由式(4-14)得:

$$\bar{X} = \frac{225 \times 1 + 312 \times 1 + 178 \times 1 + 264 \times 0 + 126 \times 0}{225 + 312 + 178 + 264 + 126} \approx 0.647 = 64.7\%$$

同意的票数占 64.7%, 超过 60%, 该项投资项目表决通过。

4. 截尾平均数

由算术平均数的计算过程可以看出,算术平均数的值会受到极端值的影响。在很多情



况下, 我们需要消除极端值对平均数的影响, 例如, 在文艺或体育比赛中, 评分要去掉若干个最高分和最低分, 再计算剩余数据的平均值, 这样计算得到的就是截尾平均数。

$$\bar{X}_\alpha = \frac{X_{(n\alpha+1)} + X_{(n\alpha+2)} + \cdots + X_{(n-n\alpha)}}{n - n\alpha} \quad (4-15)$$

式中 α ——截尾系数;

$\alpha = m/n$;

n ——数据个数;

m ——去掉的数据个数;

$X_{(1)}, X_{(2)}, \cdots, X_{(n)}$ ——将数据按升序排列后的顺序序列。

【例 4-14】 2014 年中央电视台钢琴、小提琴大赛的评分办法为: 比赛采用百分制, 按评委序号现场亮分, 每位选手的演奏得分为去掉最低分、最高分后的平均值。下列数据是 11 位评委给某一小提琴选手的分数, 按照此次大赛评分办法, 求这位选手的演奏得分。

97.2 96.7 98.3 94.8 95.9 98.1 98.7 95.7 98.2 96.8 96.3

解: 先将数据按升序排列, 求出最高分和最低分。最高分为 98.7, 最低分为 94.8。

由于 $n=11$, $m=2$, 所以 $\alpha=2/11$, 将数据代入式 (4-15) 得 $\bar{X}_{2/11}=97.025$, 即这位选手的得分是 97.025 分。

5. 算术平均数的数学性质

容易证明, 算术平均数 (均值) 有以下数学性质。

性质 1: 各变量值与其算术平均数的离差之和等于零。对于没有分组的数据有: $\sum(X-\bar{X})=0$, 对于分组的数据有: $\sum(X-\bar{X})f=0$ 。

性质 2: 各变量值与其算术平均数的离差平方和最小。对于没有分组的数据有: $\sum(X-\bar{X})^2=\min$, 对于分组的数据有: $\sum(X-\bar{X})^2f=\min$ 。

性质 1 刻画了这样一个事实: 变量值围绕着其均值为中心波动, 波动过程中形成的正的偏差与负的偏差会互相抵消。这个性质也是均值为重心的另外一种表达方式。

性质 2 刻画了这样一个事实: 变量值围绕着其均值为中心波动, 波动过程中形成的偏差的平方和为最小。这个性质为度量变量值的偏差程度奠定了基础。

这两个性质的直观表示如图 4-2 所示。

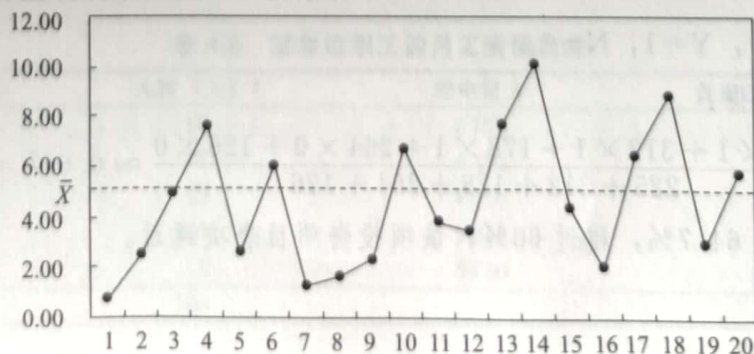


图 4-2 均值的数学性质



4.3.3 调和平均数

调和平均数是各变量值倒数的算术平均数的倒数,因而也称为倒数平均数。在实际应用中,经常会遇到已知变量值但缺少个体数量的情况,这时就要用调和平均数法计算平均指标,其计算形式有简单调和平均数和加权调和平均数两种。

1. 简单调和平均数

简单调和平均数按照调和平均数的定义是各变量值倒数的简单算术平均数的倒数。其计算公式为:

$$H = \frac{n}{\frac{1}{X_1} + \frac{1}{X_2} + \cdots + \frac{1}{X_n}} = \frac{n}{\sum_{i=1}^n \frac{1}{X_i}} \quad (4-16)$$

式中 H ——调和平均数。

【例 4-15】 一位投资者进行某只基金的定投业务,在每个月末买入 1 万元这只基金,三个月的交易记录如表 4-8 所示,如果不考虑手续费和其他费用,计算其所持基金的单位成本。

表 4-8 基金定投交易记录

月份	金额(万元)	基金单位净值(元)
1	1	1.60
2	1	1.25
3	1	2.00
合计	3	—

解: 在这个问题中,我们只知道投资者合计定投了 3 万元,但不知道共计买入的基金份额,这时需要用调和平均数来计算单位成本。

$$\text{单位成本} = \frac{\text{买入基金总金额}}{\text{买入基金总份数}} = \frac{1+1+1}{\frac{1}{1.60} + \frac{1}{1.25} + \frac{1}{2.00}} \approx 1.56(\text{元/份})$$

2. 加权调和平均数

加权调和平均数按照调和平均数的定义是各变量值倒数的加权算术平均数的倒数。其计算公式为:

$$H = \frac{m_1 + m_2 + \cdots + m_k}{\frac{m_1}{X_1} + \frac{m_2}{X_2} + \cdots + \frac{m_k}{X_k}} = \frac{\sum_{i=1}^k m_i}{\sum_{i=1}^k \frac{m_i}{X_i}} \quad (4-17)$$

式中 m ——调和平均数的权数。



【例 4-16】投资者在投资基金时，会在其价格（净值）较低时买入更多的金额或份数。表 4-9 是一位投资者买入同样基金的交易记录，如果不考虑手续费和其他费用，计算其所持基金的单位成本。

表 4-9 基金交易记录

月份	金额（万元）	基金单位净值（元）
1	2	1.60
2	3	1.25
3	1	2.00
合计	6	—

解：这个问题需要用加权调和平均数来计算单位成本。

$$\text{单位成本} = \frac{\text{买入基金总金额}}{\text{买入基金总份数}} = \frac{2+3+1}{\frac{2}{1.60} + \frac{3}{1.25} + \frac{1}{2.00}} \approx 1.51(\text{元/份})$$

从计算结果可以看到，由于这位投资者在基金价格较低时投入了较多的金额，所以拉薄了单位成本。

需要说明的是，调和平均数是各个算术平均数倒数的算术平均数的倒数，是在资料受到限制的情况下算术平均数的一种变形。

如何判断在什么情况下可以采用算术平均数或调和平均数呢？关键在于以算术平均数的基本公式为依据进行判断。当我们直接掌握了分母资料时，用算术平均数公式计算；当我们没有直接掌握分母资料而需要通过计算取得时，可考虑用调和平均数公式计算。对同一现象，计算调和平均数和算术平均数的经过是相等的，无非是因数据条件不同而采取了不同的计算形式。

如变量值有取 0 的情况，不能计算调和平均数。

4.3.4 几何平均数

几何平均数是将若干个变量值的连乘积开数次方来计算的一种平均数。它适用于对速度、比率等特殊数据计算平均数。几何平均数的计算形式分为简单几何平均数和加权几何平均数两种。

1. 简单几何平均数

简单几何平均数是 n 个变量值连乘积的 n 次方根。其计算公式为：

$$G = \sqrt[n]{X_1 \cdot X_2 \cdots X_n} = \sqrt[n]{\prod_{i=1}^n X_i} \quad (4-18)$$

式中 G ——几何平均数；

\prod ——连乘符号。



【例 4-17】 生产某产品需要经过五个流水工序完成。某日各工序制品合格率分别为 95%, 98%, 95%, 94%, 90%。要求计算该产品的平均合格率。

解: 平均合格率为:

$$G = \sqrt[5]{95\% \times 98\% \times 95\% \times 94\% \times 90\%} = 94.36\%$$

2. 加权几何平均数

对于每个变量值的次数不同的分组资料, 可采用加权几何平均数。其计算公式为:

$$G = \sqrt[f_1 + f_2 + \dots + f_m]{X_1^{f_1} \cdot X_2^{f_2} \dots X_m^{f_m}} = \sqrt[\sum_{i=1}^m f_i]{\prod_{i=1}^m X_i^{f_i}} \quad (4-19)$$

【例 4-18】 某银行的一种长期投资产品按复利计息, 各年度的利率为: 3 年为 3%, 4 年为 8%, 2 年为 10%, 3 年为 12%, 计算年平均利率。

解: 年平均利率为:

$$\begin{aligned} G &= \sqrt[\sum_{i=1}^m]{\prod_{i=1}^m x_i^{f_i}} - 1 \\ &= \sqrt[12]{1.03^3 \times 1.08^4 \times 1.10^2 \times 1.12^3} - 1 \\ &= \sqrt[12]{2.5272} - 1 = 8.03\% \end{aligned}$$

如果变量值中有零或负数, 不适合计算几何平均数。

4.3.5 众数和中位数

众数和中位数称为位置平均数, 是根据总体中处于特殊位置上的变量值来确定的代表值, 它对于整个总体来说, 具有非常直观的代表性, 因此, 常用来反映分布的集中趋势。

1. 众数

(1) 众数的含义 众数是指总体中出现次数最多的变量值, 用 M_0 表示。众数也代表分布的集中趋势, 在一些特定的情形下, 它的代表性会更好。例如, 衣服、鞋子的尺码等, 用众数代表其分布的集中趋势更加适合。

(2) 众数的计算 对于未分组数据或单项数列的数据, 可根据众数的定义确定众数, 我们只需要统计出哪一个变量值出现次数最多, 则该变量值就是众数。

【例 4-19】 一家服装专卖店统计了某天销售服装的尺码数据, 求其众数。

XL	XL	XL	XL	S	L	XL	XL	L	XL	M	M
M	M	L	XXL	M	M	XL	XL	L	L	M	L
L	M	L	L	M	L	L	XL	M	S	L	L
XL	L	L	L	M	M	S	M	L	XXL	L	XL
L	M	S	L	M	M	L	M	L	M	L	L



解：将尺寸数据进行单项式分组，结果如下：

尺码	次数
S	4
M	18
L	24
XL	12
XXL	2

由于 L 出现的次数最多，所以众数是 L。

如果数据是组距数列，需要运用公式计算众数，限于篇幅不做介绍，读者可以参阅相关资料。

众数具有以下特点。

1) 众数是以它在所有变量值中所处的位置来确定的，它不受分布数列极端值的影响，从而增强了众数对分布数列的代表性。

2) 当分组数列中没有任何一组的次数占多数，也即分布数列中没有明显的集中趋势，而是近似于均匀分布时，则该次数分配数列无众数。

3) 缺乏敏感性。这是由于众数的计算只利用了众数组及相邻组的数据信息，而算术平均数利用了全部数据信息。

2. 中位数

(1) 中位数的含义 中位数是将总体各单位的变量值按大小顺序排列起来，形成一个数列，居于数列正中间位置的变量值就是中位数，用 M_e 表示。中位数的作用与算术平均数相近，也作为所研究总体一般水平的代表值。

在数列中出现了极端变量值的情况下，用中位数作为代表值要比用算术平均数更好，因为中位数不受极端变量值的影响；如果研究目的就是为了反映中间水平，当然也应该用中位数。在统计数据的处理和分析时，可结合使用中位数。

(2) 中位数的计算 对于未分组的数据计算中位数，一般是先将变量值按大小顺序排成序列，若变量值的个数 n 为奇数，用 $(n+1) \div 2$ 的公式计算中位数所在位置，这个位置上的变量值就是中位数；若变量值的个数为偶数个，用 $\frac{n}{2}$ ， $\frac{n}{2} + 1$ 的公式计算中位数所在位置，这两个位置上的变量值的平均数就是中位数。

对于单项数列计算，中位数需要分两步进行：

第一步，使用向上累计或向下累计来确定中位数所在组的位置 $\sum f/2$ 处；

第二步，中位数所在组对应的变量值就是中位数。

【例 4-20】我们仍以表 4-4 的资料为例确定中位数的过程如表 4-10 所示。



表 4-10 新入职员工月工资累计表

按月工资分组 (元)	人数 (人)	累计次数	
		向上累计	向下累计
3 260	3	3	20
3 300	5	8	17
3 360	8	16	12
3 400	3	19	4
3 500	1	20	1
合计	20	—	—

求中位数。

解：可以按向上累计和向下累计两种方法确定中位数所在的位置。

按向上累计方法计算，中位数的位置为：

$\sum f/2 = 20/2 = 10$ ，即累计次数至 10 所在组，该例中为第三组。

则中位数 $M_e = 3\,360$ (元)。

按向下累计方法计算，中位数的位置同样为累计次数至 10 所在组，该例中其中位数也在第三组，即 $M_e = 3\,360$ (元)。

可见，不论是按向上累计方法还是按向下累计方法计算的中位数都是一样的。

中位数有一个特殊的性质：数据值与中位数之差的绝对值之和最小，即 $\sum_{i=1}^n |X_i - M_e| = \min$ 。

利用这个性质，可以在诸如配送中心的选址等问题中确定最经济的位置。

对于组距数列计算中位数的方法比较复杂，读者可以参阅其他资料。

大部分统计软件都能直接求出中位数。

特别需要说明的是，中位数不一定在所观察数据全距的中点，其位置与数据的分布有关，图 4-3 展示了不同分布对应的中位数的位置。

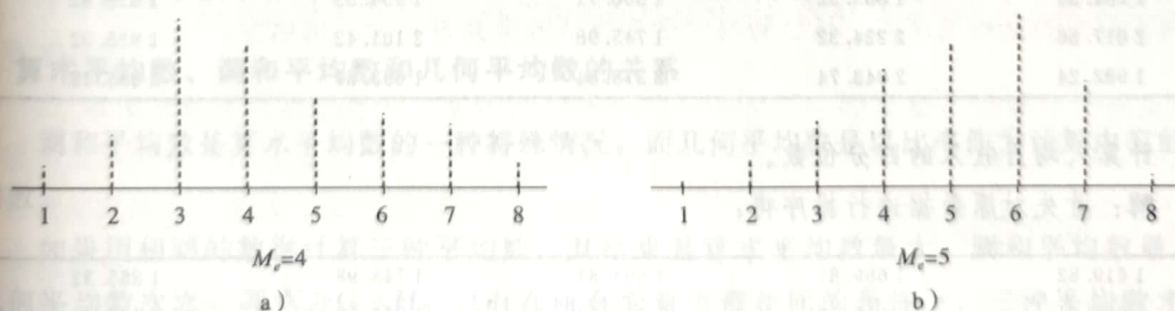


图 4-3 不同分布对应的中位数

中位数具有以下特点。

1) 中位数是以它在所有变量值中所处的位置来确定的，不受分布数列的极大或极小值影响，从而在一定程度上提高了中位数对分布数列的代表性。

2) 缺乏敏感性。这是由于中位数的计算只利用了中位数数组及部分组的数据信息，而数值平均数利用了全部数据信息。



3. 四分位数

(1) 四分位数的含义 中位数通过中间点将全部数据等分为两部分，每一部分包含全部数据的 50%。与中位数类似的还有四分位数，四分位数是通过三个点将全部数据等分四部分，每一部分包含全部数据的 25%。很显然，中间的四分位数就是中位数，因此通所说的四分位数是指处在 25% 和 75% 位置上的变量值，分别称之为下四分位数和上四分位数，记为 Q_1 和 Q_3 (Q_2 为中位数)，其含义如图 4-4 所示。

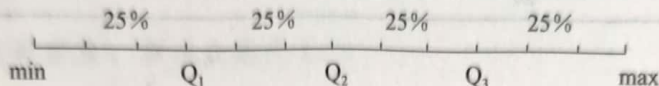


图 4-4 四分位数的含义

(2) 四分位数的计算 与中位数的计算方法类似，首先对数据进行排序，然后确定分位数所在的位置。

四分位数的位置由下式确定：

$$L_1 = (n+1) \frac{25}{100} \quad (4-20)$$

$$L_3 = (n+1) \frac{75}{100}$$

确定了四分位数的位置，位置上对应的变量值就是四分位数。当四分位数的位置不在某一个位置上时，可根据四分位数的位置，以其两侧的变量值的平均值作为四分位数。

【例 4-21】 在某城市随机抽取 20 个家庭，调查得到每个家庭的人均月收入数据如下（单位：元）。

2 084.48	1 689.81	2 167.61	2 094.80	2 289.62
1 964.50	1 865.32	1 900.71	1 954.52	1 619.82
2 017.86	2 224.32	1 745.98	2 101.42	1 915.02
1 982.24	2 043.74	2 318.34	1 693.87	2 685.45

计算人均月收入的四分位数。

解：首先对原数据进行排序得：

1 619.82	1 689.81	1 693.87	1 745.98	1 865.32
1 900.71	1 915.02	1 954.52	1 964.50	1 982.24
2 017.86	2 043.74	2 084.48	2 094.80	2 101.42
2 167.61	2 224.32	2 289.62	2 318.34	2 685.45

Q_1 的位置 $L_1 = (n+1) \frac{25}{100} = 21 \times \frac{25}{100} = 5.25$ ，即 Q_1 在第 5 个数值 (1 865.32) 和第 6 个数值 (1 900.71) 之间的位置上，因此：

$$Q_1 = (1 865.32 + 1 900.71) \div 2 = 1 883.015$$



Q_3 的位置 $L_3 = (n+1) \frac{75}{100} = 21 \times \frac{75}{100} = 15.75$ ，即 Q_3 在第 15 个数值 (2 101.42) 和第 16 个数值 (2 167.61) 之间的位置上，因此：

$$Q_3 = (2\ 101.42 + 2\ 167.61) \div 2 = 2\ 134.515$$

如果是分组数据，计算四分位数比较复杂，读者可参阅其他资料。

类似地，我们可以定义其他的分位数。

4.3.6 平均数之间的关系

1. 平均数的选择

无论是数值平均数还是位置平均数，都描述了观测数据的集中趋势，即数据在平均值附近最为集中。那么，我们应该选择哪种平均数作为集中趋势的测试值呢？

一般情况下，平均数中运用得最多的是均值、众数和中位数。均值、众数和中位数各自具有不同的特点，掌握它们之间的关系和各自的特点，有助于我们在实际应用中选择合适的测度值来描述数据的集中趋势。

均值的含义通俗易懂，直观清晰，全部数据都要参加运算，因此它是一个可靠的具有代表性的值。任何一组数据都有一个算术平均数。均值具有优良的数学性质，适合于代数方法的演算，均值是实际中应用最广泛的集中趋势测度值。但是，均值易受极端值的影响；对于偏态分布的数据，或者分散程度较高的数据均值的代表性较差。

众数是一种位置平均数，易理解，不受极端值的影响。但是众数不一定存在，它适合于存在明显集中趋势的数据，定类和定序数据的集中趋势的测度一般用众数表示。

中位数也是一种位置平均数，易理解，不受极端值的影响。中位数一定存在，一般情况下中位数介于众数和均值之间，是一个比较适中的集中趋势测度值。

2. 算术平均数、调和平均数和几何平均数的关系

调和平均数是算术平均数的一种特殊情况，而几何平均数是以比率作为计算内容的平均数。

如果用相同的数据计算三种平均数，其结果是算术平均数最大，调和平均数最小，几何平均数次之，即 $\bar{X} \geq G \geq H$ ，只有在所有变量值都相同的条件下，三种平均数才相等。但是实际应用中这样的比较没有意义，因为对于任何一个计算对象一般只适合采用一种方法来计算平均数，即不同的平均数计算方法适合不同的计算条件，必须正确地进行选择。

3. 算术平均数与众数、中位数的关系

算术平均数与中位数、众数均可代表总体的一般水平，它们之间存在着一定的数量上



的关系, 这种关系取决于数据的分布情况。

1) 如果数据分布是完全对称的钟形分布时, 算术平均数、中位数和众数三者相等, 即 $\bar{X} = M_e = M_o$, 如图 4-5 所示。

2) 如果数据分布是左偏分布时, 众数最大, 算术平均数最小, 即 $\bar{X} < M_e < M_o$, 图 4-6 所示。

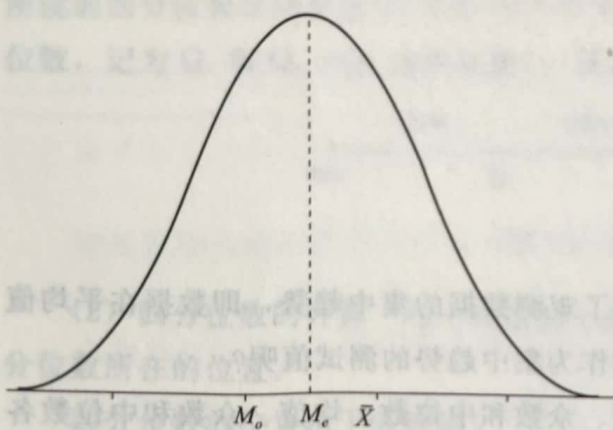


图 4-5 对称分布

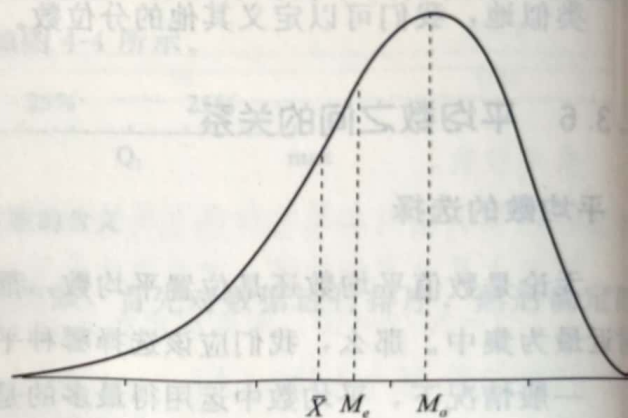


图 4-6 左偏分布

3) 如果数据分布是右偏分布时, 算术平均数最大, 众数最小, 即 $\bar{X} > M_e > M_o$, 如图 4-7 所示。

经验告诉我们, 不论是左偏分布还是右偏分布, 中位数到算术平均数的距离都大约为众数到算术平均数的距离的 1/3, 即:

$$M_e - \bar{X} \approx \frac{1}{3}(M_o - \bar{X}) \quad (4-21)$$

这个公式被称为皮尔逊经验公式。可以利用指标间的这些关系式, 从已知两个平均指标来推算另一个平均指标。

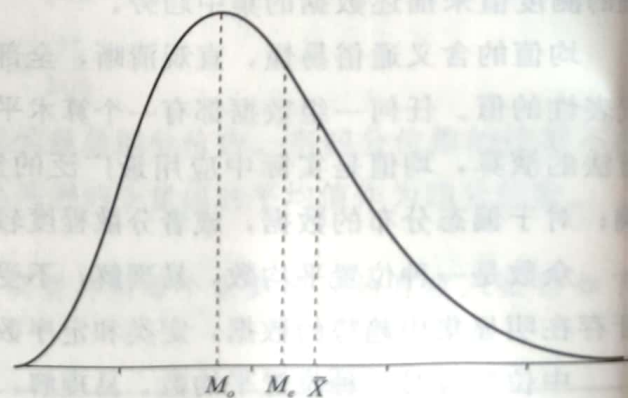


图 4-7 右偏分布

4.4 变异指标

4.4.1 变异指标的概念

平均指标描述了数据的集中趋势, 它告诉我们数据在哪个位置是集成的、密集的。但是, 仅仅依据平均指标来做判断是不可靠的, 因为我们并不知道数据的分散程度。一般来说, 数据的分散程度越大, 做判断时所要承担的风险也越大, 反之则越小。

描述数据分散程度的指标又称变异指标, 它综合反映了数据的差异性。

