

算术平均数、众数、中位数的合理运用

文 / 唐芳

随着经济社会的发展和人民文化素质的提高,我国老百姓越来越关注统计数据,统计数据是否真实、准确和科学,不仅影响我国统计数据在国际上的声誉,也影响着老百姓对统计数据的信任程度,从而影响我国社会与经济的健康稳定发展。在统计调查报告和政府权威部门公布的统计数据中,平均数是一种常见的统计数据,如平均增长速度、平均职工工资、平均价格等,用以显示社会与经济发展的一般水平和均衡状态。但由于我国统计教育还不十分普及,有些平均指标的计算方法和计算范围不够准确科学,引起了人们对平均数的质疑,从而产生了对统计数据的信任危机。如一句顺口溜就说明了平均数存在的问题:“有个暴富张千万,邻居九个穷光蛋,专家统计一平均,都是张百万”。显然,这里的平均,根本不能反映实际情况,完全失去了意义。这样的数据一出现,人们就会有被骗的感觉。其实类似情况的出现,并不是平均数本身的错,而是人们对平均数的不正确理解和不正确运用所致。作为统计教育工作者,有必要对这些问题进行深入的研究和探讨,本人在此发表自己的一些浅见。

在我国长期以来最广泛使用的是算术平均数,而多数公众由于统计知识的缺乏,也认为平均数就是算术平均数。当“平均数不代表大多数”时,自然而然就会产生对平均数的质疑。实际上,在统计中平均数不止是算术平均数,还有众数、中位数、调和平均数和几何平均数。这些平均数由于计算方法不同,具有不同的特点,反映现象特征时应用范围和代表性也就不同。只

有正确分析和掌握这些平均数的特点和应用,才能根据不同现象的分布特征选择适合的平均数,以提高平均数的准确性和可靠性。本文主要针对算术平均数、众数、中位数这三种平均数进行比较分析。

计算及特点

算术平均数是一组数据中所有数据的平均数,是将一组数据的总和除以数据个数得到的。中位数是将一组数据由小到大排序后位于中间位置的数据。若数据为奇数个,中位数为中间的那个数据;若数据为偶数个,中位数则是取中间两个数据的平均数。众数是指一组数据中出现次数最多的那个数据。下面举例说明这三种平均数的计算方法。

如为调查某单位员工收入状况,随机抽取9名员工得结果如下(元):1600、1600、1600、1800、1800、1900、1900、2500、3000,则计算三种平均数结果分别为:

$$\begin{aligned} \text{算术平均数} &= (1600+1600 \\ &+1600+1800+1800+1900+1900 \\ &+2500+3000) / 9 = 1966.67(\text{元}). \end{aligned}$$

中位数位置是第5名员工,则他的工资额1800元就是中位数。

在9个数据中,出现次数最多的是1600,所以工资额1600元就是众数。

求一组数据的平均数、众数、中位数都不是很困难,问题就出现在代表一组数据的一般水平时采用它们中的哪个来代表。特别是当这些平均数较接近时,到底用谁去代表这组数据的一般水平更合适呢?这就需要首先分析这些平均数的特点:

1. 在一组数据中平均数的个数不

同。在一组数据中,算术平均数一定有且只有一个,其数据可能等于原数据中的一个,也可能与原数据不同。而一组数据中中位数也一定有且只有一个。一组数据中众数可能有也可能没有,可能有一个也可能有多个。当一组数据中没有明显的集中趋势时,就没有众数,无法用众数代表一组数据的平均水平;当众数有多个时,可能说明总体各单位集中趋势不够明显,这时也不适合使用众数。如在上述例子中,因为9名员工工资不存在明显集中趋势,用众数1600元代表9名员工的工资水平则代表性不足。

2. 平均数的概括能力不同。算术平均数是根据一组数据中所有数据来进行计算的,受到所有信息的影响,它能够概括反映所有数据的平均水平;而中位数和众数是根据总体中处于特殊位置的个别单位或部分单位的数据来确定的,不能反映所有信息,特别是不能反映极端值的信息。因此算术平均数对数据的概括能力要比众数和中位数对数据的概括能力强,所以在传统统计分析中,人们比较习惯使用算术平均数。

3. 平均数对数据的灵敏度不同。因为算术平均数是根据数据中所有的数据来计算的,每一数据的任何变动都将在一定程度上影响到算术平均数的计算结果,而中位数和众数是根据特定位置上的数据计算的,计算时没有使用到所有的数据,因而不具有这样的特性,因此相对于中位数和众数而言算术平均数对一组数据表现出更大的灵敏度。如果数据中出现个别极端值,对算术平均数而言,会拉高或拉低平均水平,从而出现前面讲到的“张百万”

的情况出现,影响了平均数的代表性。在上述例子中,算术平均数1966.67元受到了极端值3000元的影响,使平均数偏高,在9名员工中只有两名员工的工资达到平均水平,显然这时的算术平均数对9名员工工资水平的代表性较差。而中位数和众数则不受极端值的影响,避免了个别数字引起的误解。前面提到众数对平均水平的代表性也不足,这里用中位数1800元来代表比较合适。

4. 平均数的适用范围不同。一般来说,由于算术平均数对于数据的量化尺度要求较高,算术平均数比中位数众数的适用范围要窄,只适用于定距尺度和定比尺度的数据;中位数不仅适用于定距尺度和定比尺度的数据,还适用于定序尺度的数据;而众数的适用范围更广,不仅适用于上述三种数据,还适用各种定类尺度的数据。因而在一些无法适用算术平均数的场合,众数和中位数不失为一种独特且有用的统计分析工具。

通过上述分析我们可以看出,算术平均数、众数、中位数的定义不同,它们具有不同的特点和不同的适用范围,在使用时应该正确恰当地根据不同的情况选择不同的平均数,这样才能够提高这些平均数的代表性,更客观准确地反映和描述事物现象的本质特征。

在经济生活中的合理运用

1. 分析定类数据的平均水平时,宜采用众数。从统计学而言,理论上,一切认识的对象均可被量化。而其量化的方法则无外乎四种——定比、定距、定序、定类。定类尺度只进行分类或分组,不反映种类的优劣、量的大小或顺序,不可以进行数学运算,例如人口分为男女两类。定序尺度可以反映种类的优劣、量的大小或顺序,不可以进行数学运算,例如将居民分为高、中、低三类。定距尺度可以现象在量方面的精确差异,可以进行加减运算,例如温度。定比尺度除有上述三种尺度的全部性质之外,还可以测量不同数据之间的比例或比率关系,可以进行加减乘除运算,例如收入、价格、销售量。

众数作为位置平均数能够适用于所有的数据类型,尤其是定类数据因为不能

计算中位数或算术平均数,它的中等水平或一般水平只能由众数来代表,在很多实际问题中得到应用。诸如商品销售情况的统计中,作为最受欢迎的、最大众化的、最普遍的、销售最大的类型;在民意测验中,作为反映民众倾向的意见等等。

2. 现象分布比较均衡时,宜采用算术平均数。现实生活中,大多数现象的分布是钟型分布,显示出国民经济运行过程中的均衡状态,而算术平均数是所有数据的平均水平,能反映所有数据信息,加之这种方法简单易懂,符合代数运算,所以在数据分布比较均匀的情况下,能真实准确地反映现象的一般水平和本质特征。比如,在我国实行计划经济时代,职工收入分配比较均匀,差距较小,这时用算术平均数来反映我国职工收入状况比较适宜。但在我国实行市场经济的今天,经济形势急剧变化,社会群体迅速分化,地区差异拉大,这时人们的收入分配十分悬殊,分配差距相当地大,这时用算术平均数作为人们收入状况的代表值,则其代表性不强,说服力不够。

3. 当数据中出现极端值时,宜采用中位数。据统计资料显示,目前全国城市居民收入最低的五分之一人口只拥有全部收入的2.75%,仅为收入最高的五分之一人口所拥有收入的4.6%。各占总人口20%的最高和最低收入两大群体,收入差距已达33倍。在我国社会贫富悬殊拉大的今天,在计算反映人民生活水平的平均数时,如果再用算术平均数作为这些数据的代表值,会使平均数出现虚高或虚低,使人们对客观事实缺乏正确的认识。中位数是总体中所有变量值按顺序排列后处在中间位置的那个变量值,它不受极端值即极大值和极小值的影响,最适合反映数量差异悬殊现象的数量特点,所以在反映人民生活水平的一般水平时,应使用中位数来计算,如人均收入、人均住房面积。在医学卫生统计中,世界各国在反映婴儿每个月的标准身高和体重以及血液中各微量元素含量均使用的是中位数。

4. 现象存在明显集中趋势时,宜采用众数。平均数是反映现象分布集中趋势的指标,反映大部分单位的一般水

平,而众数是一组数据中出现次数最多的数据,在现象分布存在明显集中趋势时,众数作为平均水平的代表性较高。但与算术平均数相比,有关众数的介绍和研究较少,人们对它的关注度不高,实际上众数的适用范围比算术平均数更广,如前面提到的定类尺度就适用众数代为代表值。众数的应用还常出现在不确定情况下理性决策中。例如,成衣、鞋袜、帽子、自行车等商品的生产 and 销售,厂家或商店为了了解消费者的需求,所关心的不是这些商品的号码、尺寸或型号的算术平均数,而且也无法掌握所有需求者的信息,这时采用销售出去的商品的号码、尺寸、型号的众数作为平均数,更具代表性也更为现实。又如,在竞选中,候选人为了获胜,一般是选择在选民最感兴趣话题上作文章。

5. 将平均数与其他指标相结合。有统计常识的人都知道,平均数是一组数据的合计数除以个数,这组数据的差异越大,平均数的代表性也就越小。现阶段的中国,人口众多,地区发展不平衡,还处在底层大、中间小的“金字塔型”,绝大多数财富聚集在处于塔尖的少数人。“金字塔型”的中国如果凡事都参考平均数,无异于利用“障眼法”,忽略贫富差距,掩盖社会矛盾,助长“马太效应”。因而统计上的平均数虽然仍有其作用和价值,但其对社会差距和矛盾的掩饰,必然会引起公众的反感和不满。在贫富差距扩大、基尼系数已经超过警戒线的今天,许多情况不宜只简单地公布一个笼统的平均数,不妨多公布一些能够表现差异和差距的平均数,如高收入的平均数、低收入的平均数等,或者是公布一些反映差距的指标值,如标准差、离散系数。这样的统计数据能更全面地反映现象的真实面目,从而取信于民。

在平均数受到广泛质疑的今天,我们不能因为质疑而否认这些指标使用的意义,而应该分析问题出现的原因和解决办法,以更好地运用这些指标,更好地发挥它们在社会经济生活中的作用。

(作者单位:广东顺德职业技术学院)