

【经济管理研究】

关于统计学基础中几个问题的探讨

刘卫东^①

(辽东学院 经济学院, 辽宁 丹东 118001)

摘 要: 在统计学基础教学中, 会遇到一些问题, 这些问题中, 有的是概念表述不准确, 有些是内容表述不透彻。这些问题的存在, 使学生的学习过程中, 会产生一些疑惑。针对总体与总体单位之间的关系, 变量的概念, 算术平均数与强度相对数的区别, 区间估计问题以及描述统计与推断统计的区别问题, 进行了探讨, 以期起到抛砖引玉的作用。

关键词: 统计学; 教学内容; 变量; 强度相对数

中图分类号: F222

文献标志码: A

文章编号: 1672-8572 (2014) 05-0104-04

一、关于统计总体和总体单位的关系问题

统计总体一般定义为: 在同一性质基础上结合起来的许多个别单位的整体。总体单位是构成总体的每一个单位。从上述定义中可以看出, 统计总体有一个特点就是一个总体中, 必须有许多个别单位, 也就是我们通常所说的总体的大量性。但有些教材在说明总体和总体单位的关系时, 会提出一种观点: 统计总体和总体单位在不同的条件下可以互换, 即在一定的条件下, 某一现象, 可以作为总体, 而在另一条件下, 则可以成为总体单位。例如, 在一些教材中这样描述: 当研究 A 市民营企业的生产经营情况时, 该市的全部民营企业就是一个总体, 其中每一家民营企业就是一个总体单位; 如果要研究一家民营企业的生产经营情况, 那么这家民营企业就成为总体了。

笔者认为, 这种说法是不妥当的。从总体的定义上可以看出, 总体必须由许多单位组成, 一个单位是不能构成一个总体的, 上例中表明, 在研究一家民营企业的生产经营状况时, 该家民营企业即为总体, 这表明一家企业也可以作为一个总体, 不符合对总体的定义。如果说研究这家民营企业的所有职工情况, 那么, 此时这家企业的所有职工才是总体, 而这家企业不能成为总体, 如果说这家企业

是总体, 那么其总体单位是什么呢? 所以笔者认为, 总体和总体单位在不同的情况下可以互换的表述是不恰当的。

拓展问题: 标志和指标互换问题

在一些教材中, 由于存在着总体和总体单位互换的观点, 从而得出了标志和指标也可以互换的结论。因为指标是说明总体的, 而标志是说明总体单位的, 既然总体变成了标志, 那么原来说明总体的指标, 就变成了说明总体单位的了, 也就变成了标志了。例如, 有的教材中的例子提到: 我们研究一个工业企业(总体)的生产情况时, 该企业的工业总产值、职工人数等是反映该企业的生产情况的统计指标; 而当我们研究全国或一个地区的工业生产情况时, 每个企业(总体单位)的工业总产值、职工人数等, 则变成了反映企业特征的标志。笔者认为, 这种提法是有问题的。前文中已经提出, 总体和总体单位是不能互换的, 在这一前提下, 标志和指标当然也就无法互换。该例中的一个企业是不能成为总体的, 既然一个企业不能成为总体, 那么反映这一个企业的工业总产值、职工人数, 也只能是标志, 而不可能成为指标。所以指标和标志互换的提法是不正确的。

^① 收稿日期: 2014-07-20

作者简介: 刘卫东(1964—), 男, 沈阳人, 硕士, 副教授, 研究方向: 统计学。

二、关于变量的概念

统计学教材中,通常把变量定义为:可变的数量标志。这种说法是否合适,还需要进一步探讨。

变量,顾名思义,是一个变化的量,也就是其数值不确定的量。如果将变量定义为可变的数量标志,那么指标也就不是变量,因为从统计学中的概念可知,标志是说明总体单位的,而指标是说明总体的。但在时间序列分析时,时间序列的构成要素包括两个因素,一个是现象所属的时间,一个是现象的指标数值。那如果将某个企业不同时期的人员加以平均的话,此时的企业人员数是变量还是指标呢,按指标的定义:指标是描述总体特征的,而说明一个企业不同时期的人员情况,这个时候的企业仅是一个总体单位,是不能构成总体的。所以这个时候的企业人员数更应该是一个变量,而不是指标。因此,如果只将变量定义为可变的数量标志,那对于这种情况就无法解释了。

笔者认为:变量可以定义为一个变化的量,无论是标志也好,指标也好,都可以定义为变量,这样不仅可以使问题简单化,而且更便于理解和应用。

拓展问题:关于连续变量与离散变量的问题

在解释连续变量与离散变量时,有这样的描述:离散变量是用整数表示的,而连续变量可以用小数表示。这里的问题是,整数还是小数的计量单位是什么?比如人数,我们认为其是离散变量,但如果我们的计量单位是百人、千人的话,那么人数也可以表示为0.25千人,这样也出现了小数,是不是说其就可以作为连续变量了呢?再如,企业利润这个变量,是连续的呢,还是离散的呢?我们认为,是连续的,但如果精确到分的计量单位,这个利润变量还会出现小数吗?所以笔者认为,在介绍连续变量和离散变量时,除了介绍其概念以外,还需要在变量的计量单位上进行说明,以便更好地理解变量的类型。

三、关于抽样推断的区间估计问题

许多统计学教材中,在介绍抽样推断的区间估计内容时,都将参数的估计区间设定为“平均数的区间估计”和“成数的区间估计”,其中为极限

抽样误差,为概率度。这里的关键问题是概率度的确定方法。在许多统计学教材中,概率度是根据事先给定的估计的置信程度,查正态分布概率表来确定的。这里就存在了一个问题,那就是在计算概率度时,并不是任何情况都要查正态分布概率表来获得概率度的数值,而是需要分别对待的。正确的说法是,在正态分布下,方差已知或方差未知但却为大样本的情况下,可以查正态分布概率表来获得概率度的数值,而在方差未知,且为小样本的情况下,我们则需要查分布表来获得概率度的数值。这在教学过程中,是应该明确说明的。

四、关于算术平均数和强度相对数的区别

在统计教学中,有些问题比较难以理解和区别,其中算术平均数和强度相对数区别就很难理解。有些强度相对数有较强的“平均”意思,如“人均钢产量”,而平均指标也具有平均的意思,如“职工人均工资”。如何区别算术平均数和强度相对数呢?

首先,从概念上看,二者是不同的。强度相对数是两个有联系而性质不同的总体对比而形成相对数,反映两个不同总体现象对比所形成的密度、强度;算术平均数是反映同质总体单位标志值一般水平的指标,即分子和分母同属于一个总体,一个是总体的标志总量,一个是总体的单位总量。从计算公式上看,二者也是不相同的,算术平均数分子、分母分别是同一总体的总体标志总量和总体单位总量,是总体标志总量除以总体单位总量,分子、分母具有一一对应的关系,即分母每一个总体单位都在分子可找到与之对应的标志值,反之,分子每一个标志值都可以在分母中找到与之对应的总体单位。例如,计算企业职工的平均工资,是企业的工资总额除以企业的职工人数,这里每个职工都具有工资这一标志,也即都有其相应的工资额,分母与分子具有一一对应关系;而强度相对数是两个总体现象之比,分子分母没有一一对应关系。例如,人均钢产量,分子是钢产量,而分母是人数,但分母每个人并不都具有钢产量这一标志,许多人是不会生产出钢的,也就是分子和分母没有一一对应关系。

五、描述统计和推断统计问题

统计学中,将统计分析方法分为描述统计和推断统计。在某些教材中,对描述统计和推断统计是这样定义的:描述统计是统计分析的最基本内容,是指应用统计指标、统计表、统计图等方法,对资料的数量特征及其分布规律进行测定和描述;而统计推断是指通过抽样方式,利用样本估计总体特征的过程,包括参数估计和假设检验两项内容^[1]。

笔者认为:描述统计应该定义为对历史数据的描述和分析,是对有关事物进行准确的表述,这种描述和分析,即可以用文字描述,也可以用图表来展示,更多的是用指标来体现;而推断统计是利用历史资料去推断未来情况或用部分单位的资料去推断未知总体情况的分析方法,是根据所掌握的历史资料去对未知的事物进行的一种估计和判断,不是对未知事物的结果做出的准确性判断,而应该是具有一定把握的概率估计。从这个意义上讲,不仅参数估计和假设检验属于推断统计的内容,同样回归分析、时间序列的预测分析等一些具有预测和估计的统计分析,也应该属于推断统计的内容。

六、样本容量与样本个数问题

样本容量也称样本单位数或样本量,是从总体按随机原则抽取的部分单位数量。例如,从总体10 000个单位中,按随机原则抽取100个单位组成样本,其样本容量即为100。

样本个数也称样本的可能数目,是指按照随机原则从总体中抽取样本单位组成样本的所有可能的组合数。从一个总体中抽取样本单位,会有许多种组合,有多少种组合,就会有多少样本可能数目,样本个数的多少,与样本容量以及抽样方法等因素都有直接关系,是一个比较复杂的问题。例如,总体单位数为4,样本容量为2,如果按不重复抽样的方式抽取样本,则可能的样本个数为6个,假设这4个总体单位为A、B、C、D,则这6个样本分别为:AB、AC、AD、BC、BD、CD。

在统计教学中,经常混淆样本个数与样本容量的概念。这种概念上的混淆,是一种原则上的错误,会直接影响到统计推断的一些其他问题,如对

抽样平均误差概念的理解等问题,因此在教学中,要明确两者的不同。

七、标准差与抽样平均误差、估计标准误差的关系问题

标准差是标志变异指标的一种,是用来衡量变量值差异程度的指标。它是各总体单位的标志值与其平均数离差平方的算术平均数的平方根。

抽样平均误差是所有可能的样本指标与总体指标的平均离差,是样本指标的标准差。

估计标准误差是说明实际值与其估计值之间相对偏离程度指标,主要用来衡量回归方程的代表性。

从根本上说,三者的本质是一样的,都是衡量差异程度的指标,都是标准差,三个指标都可以用来衡量平均数或估计值的代表性大小。所不同的是,标准差是用来衡量所有变量值差异程度的,可以反映平均数的代表性大小;抽样平均误差是用来衡量样本指标的差异程度的,可以衡量抽样推断的精确程度;而估计标准误差是用来衡量实际值与估计值偏离程度的,可以用来衡量理论值对实际值的代表程度。

三个指标在计算公式上,也略有不同。标准差在计算公式的根号下的分母为总体单位的个数;而抽样平均误差计算公式中的根号下的分母为样本单位的个数减1,是因为在计算其平均数时,失去了一个自由度;估计标准误差的计算公式中的根号下的分母为观测值个数减2,是因为其在计算回归方程(一元回归方程)的参数时,失去了两个自由度。

八、统计教学手段问题

统计学是研究现象数量方面的科学,涉及大量的计算问题。传统的统计教学,都是在黑板上或屏幕上演示,效率低,效果差。随着计算机的广泛应用,统计分析软件的不断发展,教学设施越来越完善,教学手段越来越丰富,这为统计教学采取更加有效的教学手段,尤其是利用计算机及其软件方面奠定了基础。在众多的统计计算分析工具中,Excel是一种非常简单有效的工具。Excel提供了超强的统计分析程序,范围涵盖了最基本的统计分析^[2]。利用“分析工具库”可以用来进行统计绘

图、制表、描述统计量计算、参数估计、假设检验、方差分析、相关回归分析、时间序列分析、抽样、数据变换；通过 78 个统计函数，也可以实现许多统计计算。

笔者认为：统计学的教学由于其计算量较大，其教学既可以在安装统计软件的实验室里授课，也可以利用多媒体进行教学，在教学过程中，对于相关的统计计算，可以利用 Excel 进行分析计算，这样不仅可以节省授课时间，扩大课堂信息量，更能充分利用现代化设备，保证计算的准确性及便利性，提高授课效率。

九、方式、方法问题

在统计学中，经常会提到方式和方法问题。从字面意义上来看，二者没有什么太大区别，是近义词，但在统计中，方式和方法中还是有很大区别的。

在统计学中，提到方式和方法的内容主要是：调查方式和方法，抽样方式和方法。

首先来讨论调查方式和方法。调查方式在统计学中，指的是统计调查的组织方式，主要包括统计报表、普查、抽样调查、重点调查和典型调查；而调查方法则指的是调查的技术方法，包括询问调查法、观察法、实验法。调查方式是组织调查的方式，而调查方法是调查的技术方法，是收集资料的具体方法。

其次来讨论抽样方式和抽样方法。抽样方式是指抽取样本单位的方式，即通过什么方式来构成样本，其方法主要包括简单随机抽样、类型抽样、机械抽样、整群抽样和多阶段抽样，其主要特征是是否对总体进行相关的处理，然后抽取样本单位。如在分类抽样方式下，先对总体进行分类，然后再从不同类别中抽取单位共同组成样本。抽样的方法是如何从总体中抽样样本单位的方法，包括重复抽样和不重复抽样，其特点是对已经抽中的样本单位是否放回总体中参加下一次抽取。

参考文献：

- [1] 袁剑秋. 基于关联规则算法在数据挖掘中的研究与应用 [D]. 成都：成都理工大学，2009.
- [2] 牛永鑫，王新. Excel 在仪器标准曲线绘制中的应用 [J]. 科技资讯，2008（13）：238 - 239.

（责任编辑：游媛媛）

Some Issues in Fundamental Statistics

LIU Wei - dong

(College of Economics, Eastern Liaoning University, Dandong 118001, China)

Abstract: Some issues such as the relation between totality and totality unit, the concept of variable, the difference between arithmetic average and strength relative number, interval estimation, and the difference between descriptive statistics and inferential statistics are discussed.

Key words: statistics; teaching content; variable; strength relative number