

高校大学生学风调查的抽样设计

陈家作 北方工业大学经济管理学院数量经济学系
张加强 北方工业大学理学院
侯志强 北方工业大学统计系

[摘要]利用对前两级单元均进行分层时的自加权三阶段抽样设计来对高校大学生学风进行抽样调查,不仅可以弥补以往抽样调查的不足,而且总体总量的估计量形式简单。

[关键词]自加权三阶段抽样设计 学风

作者简介:陈家作(1983-),男,广西柳州人,北方工业大学经济管理学院数量经济学系,在校研究生;张加强(1972-),男,山西长治人,北方工业大学理学院,助理研究员,主要从事思想政治教育研究;侯志强(1970-),男,山西曲沃人,北方工业大学统计系,副教授,主要从事抽样技术研究。

一、引言

为落实科教兴国战略,我国高等教育迅猛发展,正在逐步实现高等教育大众化的目标。教育部副部长赵沁平在全国高校学风建设会上强调指出:“学风建设是高等教育改革和发展中一个永恒的主题。优良的学风是保证提高教育质量的重要条件。因此,高校应把学风建设作为提高办学质量的头等大事来抓。”

不少学者已经对高校学风进行了研究和探讨。河北理工大学冶金与能源学院为了全面了解该校学生的学习状况,于2006年6月开展了一次抽样调查。调查涉及到该院3个系、4个专业的1100余名在校生。二年级学生占34.13%,三年级学生占35.79%。沈阳农业大学园艺学院为了深刻地研究学生学习目的和学习动机方面存在的问题,共下发问卷200份,其中99级40份,2000级60份,2001级100份。中国地质大学以某教育部直属高校的全日制在校本科大学生为对象,在全校学生中按系科、专业和年级进行分层抽样,涉及到12个院系、30多个专业的在校生。

可以看到,上述的研究中并没有很好地设计抽样方案,并且没有进行精度分析,甚至可以说这些抽样都是非随机的,也就是非概率意义下的抽样。信度值得质疑。

二、对前两级单元均进行分层时的自加权三阶段抽样设计

大型抽样调查总是采用分层多阶段抽样。分层多阶段抽样若采用自加权的抽样设计,则总体总量的估计量形式简单,易于计算。根据高校大学生学风调查的实际状况与需要,本文将采用对前两级单元均进行分层时的自加权三阶段抽样设计,并考虑抽样精度问题。

自加权的分层多阶段抽样设计要求最后一阶采用放回或不放回的简单随机抽样,而当各最终抽样单元规模差异较大时,简单随机抽样估计量的方差偏大。为了提高精度,通常将最终抽样单元按照规模大致相同来设置。在各阶段分配样本量时,首先确定最后一阶中每个班级分配的样本量。比如说总样本量为100,确定每个班级分配5个样本,那么就需要抽取20个班级。然后再根据比例将班级数量分配到第二阶,依次类推。与此同时,需要考虑的是调查费用的问题。

对于分层抽样来说,层间方差越大越好。尤其是多阶段分层抽样,第一阶段的分层效果基本决定了整个抽样设计的精度。经过观察,发现不同的学院和年级之间同学们对学习态度的有着很大的差异。所以在第一阶段时,将学院和年级作为分层的标志。在学院与年级之下

为专业,每个专业有着不同数量的班级,共同构成第二阶段的抽样单元。第三阶段抽样单元为抽中班级的所有同学。

第一、二阶段可以采用简单随机抽样或者系统抽样。相比之下,系统抽样更加简单易行。但是由于在抽样框完备并且样本量相同情况下,简单随机抽样的精度更高,所以还是采用简单随机抽样。第三阶段必须采用放回或不放回的简单随机抽样。不放回的简单随机抽样精度更高,实施起来也比较容易,所以是我们的首选。同理,在第一、二阶段也应采用不放回的简单随机抽样。

在抽取样本时,可以使用SPSS或者EXCEL软件中的随机数发生器来实现。在第三阶段的抽样中,由于被抽中班级的人数不一定相等,所以应该用软件对每个班级都进行抽取;同理,第二阶段的专业与班级也不完全一样,所以都要单独抽取。

在第一阶段抽取 n 个一级单元;在第二阶段抽样中,在每个中选的一级单元内均抽取 m 个二级单元;最后在每个中选的二级单元内均抽取 k 个三级单元。前两阶段按照各层所包含的三级单元数量比例分配样本量,第三阶段抽取三级单元时采用不放回简单随机抽样。这样做保证了该抽样设计为对前两级单元均进行分层时的自加权三阶段抽样设计。

根据自加权的方差以及相对误差的计算公式,同时规定相对误差不得超过一定数值,即提出精度的要求,有了总体的方差或其估计量后就可以算出所需要的总样本量 A ,且应该有 $A=n \times m \times k$ 成立。在实际工作中,计算得到的总样本量不一定能满足被整除的性质,同时不能为了总样本量被整除而缩小总样本量,从而导致精度不高。故应该适当地放大样本量。

三、自加权三阶段抽样设计的其他问题

(一) 抽样框问题

在高校大学生学风调查中,一阶抽样框由高校所有学院和年级构成。由于学院与年级是确定的,天然生成了完备的抽样框,一阶段的抽样框问题迎刃而解。需要注意的是某些专业比较特别,例如建筑学为五年制,所以在构造抽样框时要加以考虑,不能忽略。二阶抽样框由专业与班级构成,三阶抽样框由某个班级的所有同学组成。所以说在高校的环境下抽样框是天然与完备的。

(二) 样本量问题

在没有其他可以借鉴的经验条件下,可以首先进行预调查。样本量的确定可以采用美国的经验样本量确定方法,即样本容量为问卷问题数的50倍。例如,所设计的问卷包括15个问题,那么所需要的样本容量为 $50 \times$

$15=750$ 。在得到数据后,根据上述方差以及相对误差计算公式来计算方差以及相对误差,评判是否符合精度的要求。如果不符合,根据调查得到的方差数值以及精度要求逆推得到样本容量。进行二次抽样,补足容量。

(三) 核心指标问题

在高校学风评价指标体系中,包含了诸多变量。但是在设计抽样方案的时候不可能把所有指标全部考虑到,不可能对所有指标都进行精度分析。所以只能选取某个核心指标,根据其方差与相对误差来确定样本容量,控制调查精度。在高校学风评价指标体系中适当地选取一个最有代表性的指标有助于提高抽样的准确度以及提供更多的方便。学风的好坏应该首先体现在学习成绩上,而学习成绩是个连续型变量,能够客观地反映大学学生的学习情况,所以核心指标应该是大学学生的学习成绩。

(四) 无回答问题

采用学习成绩作为核心指标来做总体以及均值估计,并计算抽样精度是行之有效的。如果由于特殊原因,比如被抽中对象已经休学等,未能及时地获取样本信息,则应该对缺失值进行处理。据观察,学习成绩的方差是比较大的。如果采用前后平均或者总体平均来替代缺失值,必然导致非抽样误差的增加。重新抽样是比较可靠的办法。

(五) 敏感性问题

高校学风调查肯定会涉及到一定比例的敏感性问题,比如是否受过处分,学习成绩是多少等。如果直接采用问卷调查的形式可能会造成大量的无回答,使得抽样效率低下。所以对于敏感性问题,应该事先设计一整套敏感性问题调查方法来避免无回答误差。

四、结束语

以往的学风调查都没有经过严格的抽样设计和精度分析,可信度不高。应用对前两级单元均进行分层时的自加权三阶段抽样设计来调查高校学风,可以经过简单的计算得到估计量与调查精度。在调查费用可允许的范围内,通过控制样本量使调查精度在可接受的区间里,提高可信度,为后续的学风整改提供依据。

参考文献

- [1]龙跃、唐国章、齐立强,大学生学风状况调查及分析[J],河北理工大学学报,2007(5):95-97
- [2]田爱民、李贺,大学生学风状况调查及启示[J],求实,2003(s2):214-215