The Discussion on Sampling Capacity by Variable

Hu Xiaoli, Wu Wenling

Nanjing College of Information and Technology, Nanjing 210046

Abstract

This paper discussed the sampling capacity in two different methods. The first method uses $(1.96s/d)^2$ to estimate. The second method is in GB/T 8054-2008. The two methods always give two different sampling sizes. In order to compare the two methods, this paper discussed the variety of sampling capacity and the different mean value and standard deviation. At last, this paper suggested to the maximum of sampling size is 60 by variable.

Keywords: Hypothesis Testing, Sampling Survey, Sampling Capacity

计量型调查中样本量的讨论

胡小丽 吴文伶

南京信息职业技术学院, 江苏 南京 210046

摘要: 抽样调查是一种常用的调查方式,但对于计量类值的调查,在经验样本量估计值和我国标准 GB/T8054 中的推荐样本量之间存在着一些差异。论文以大学生群体的身高和体重为调查对象,对有较大差异的抽样量进行了比较,利用数理统计方法,从均值和标准偏差的角度来判断两者与总体之间的差异显著性。从而给在计量值的抽样调查中的抽样量建议。本次调查的对象选择为大学生群体的体重和身高,其数据服从正态分布,结果具有可比性,研究结果也可以拓展在其他服从正态分布的计量值抽样调查。

关键词: 假设建议; 抽样调查; 样本量

引言

抽样调查是一种常用的调查方式,使用非常广泛因为从时间和金钱角度来考虑这种调查方式既快捷又经济。抽样调查的一个关键问题就是样本量的科学确定。样本量的大小直接影响抽样误差、调查的费用、调查所需的时间、调查访员的数量以及其他一些重要的现场操作的限制条件^[1]。样本量过大,会造成人力、物力和财力的浪费;样本量过小,会造成抽样误差增大,影响抽样推断的可靠程度^[2]。

关于抽样调查中样本量的确定,通常受到研究经费及调查时间的限制^[3],国内外学者在定量两个方面研究可以分为理论计算和经验判断两种类型。对于样本量的理论计算,主要根据指标估计量的方差公式和费用函数求解最优样本量对于给定的费用,使估计量的方差达到最小,或者对于给定的估计量方差,使得总费用达到最小的样本量。本文主要定理研究问题中如果样本量才更优。

1 问题提出和解决

在定量值的抽样调查方面,社会调查经验样本量计算出推荐值 $(n=1.96\sigma/d^2)$,与标准 GB/T8054-2008 中的推荐样本量存在着一些差异。通常前者计算出来的值比后者查表所得的值要大。而多数人认为,样本量越大越能反映实际情况,这种观点是不全面的。

本文以大学生群体为案例,拟通过实际案例,对这种样本量的差异性带来的均值和标准差的异同,进行比较。通过不同的抽样量分布计算其对应的均值和标准差,与设定总体进行假设检验,比较异同。

为了更明确地比较抽样量异同带来的结果差异，本文选择的调查目标为服从正态分布的大学生群体的身高、体重。且只讨论样本量的异同对结果的影响问题，而不讨论如整群抽样、分层抽样或系统抽样等抽样方法的问题。未经说明本文采用的是 GB/T10111 阐述的简单随机抽样。

1.1 抽样量选取的两种方法

经验样本量值估算方法是：设 μ 为真值， $\bar{x}$ 为观察值，则 μ 与 $\bar{x}$ 不会完全相同，会存在一定的偏差，并且在不同的调查中这种偏差也会不同，因此可以将这种偏差看成是一个随机变量^[2]。如果给出一个允许的最大偏差 d ，且要求 $|\bar{x} - \mu| < d$ 发生的可能性较大。通常要求达到 $1-\alpha$ ，即要求 $P(|\bar{x} - \mu| < d) = 1-\alpha$ ，那么可以用 $n=(\mu_{1-\alpha/2}\sigma/d)^2 \dots\dots (1)$

来估计样本量的大小。一般取^[5] $\alpha=0.05$ ，则公式可以简化为： $n=(1.96\sigma/d)^2$ ，此时的 σ 可以根据样本的实际情况进行估算^[4]，用 s 的值来作为 $\hat{\sigma}$ 。

而根据标准 GB/T8054 的公式计算，其样本量的要求与实际群体的要求相比较而言较高。

在 GB/T8054 中，抽样量的获取是根据 $B=(\mu_{1U}-\mu_{0U})/S$ 或 $B'=(\mu_{0L}-\mu_{1L})/S$ 的计算值，可由 GB/T 8054 中的表 4 查得。其中 μ_{1U} 表示：极限质量的上规格限， μ_{0U} 表示合格质量的上规格限； μ_{1L} 表示极限质量的下规格限， μ_{0L} 表示合格质量的下规格限。由计算得出的 B 值，根据 GB/T 8054 表 4 查得抽样量推荐值。根据公式特点，在运用 GB/T8054 的方法取样本量时，均值的差异几乎不会对对样本量有影响，但标准差的大小与可接收误差范围的比值将直接影响样本量。

下表 1 将在依次改变误差的大小，用经验样本量估计公式 ($n=(1.96s/d)^2$) 对样本量进行估算；同时也依次改变 B 的大小，依次查表得到新的抽样量。将两者所得样本量大小的改变进行动态比较。

研究数据来源是根据校外调查的大学生身高、体重。单位分别是厘米和公斤。用这个数据来检验是以为其服从正态分布，其数据可根据现有的假设检验进行判断。从而比较两种抽样量对结果的影响是否显著。

调查的总体有效样本数量为 1150，选用此样本的均值和标准差来对比样本量改变时，均值和标准差的差异显著性程度，由此对比样本量的改变对调查结果的影响显著性差异。

在 GB/T8054 中对抽样量的取值方法是通过查表取得。但无论是 σ 已知条件下查表 2，还是 σ 未知条件下，查表 4 其最大取得抽样量均不超过 60^[5]。而根据社会调查的抽样量估计公式，其样本量的大小与误差的平方成反比。当 B 与 d 对应等量误差时，标准 GB/T8054 推荐样本量与公式 (1) 估计样本量也并不相同。尤其是 d 取得比较小的时候 (d 是标准差的十分之一以内)，两种方法估计的样本量相差比较大。

下表 1 是根据实际抽样调查的结果，进行了取不同 B 值^[6]和 d 值条件下抽样量的比较，列出了样本量改变的明细。将两种方法估算的样本量的变化用数字展现出来。表中同一序号所对应的 B 值 和 d 值在置信度相同的情况下，按接受概率来用 $1-\alpha$ ，其误差范围也是一致的。在此条件下比较就非常容易比较两种方法估算的抽样量的异同。

表 1: B 值和 d 值改变抽样量的变化

序号	μ_{1L}	μ_{0L}	μ_{0u}	μ_{1u}	B	n_1	d	n_2
1	162	166	168	172	0.5	36	1	256
2	161	166	168	173	0.625	24	2	64
3	160	166	168	174	0.75	18	3	29
4	159	166	168	175	0.875	14	4	16
5	158	166	168	176	1	11	5	11
6	157	166	168	177	1.125	9	6	8

1.2 抽样量的改变对结果的影响

在误差 d 较小的时候，约是标准差的十分之一左右，两种样本量估计值的差异是最大的。本文根据这种差异来进行统计学意义上的对比计算。上表 1 中，序号 1 对应的等量大小的误差，根据经验样本量的公式估算抽样量 $n_1=256$ ，根据 GB/T8054 查表所得的抽样量 $n_2=36$ 。当误差的变化达到一定量时，即时标准差的 1 倍时，两种方法所取得的样本量一样，都是 11。我们将之分别与 $n=1150$ 时的均值和标准差做假设检验。需要说明的是，在比较过程中，取的 $n_1=256$ 个值采用的是整群抽样。取得的 $n_1=36$ 和 $n_3=11$ 个值，采用的是简单随机抽样。下表 2 是在取三个样本量的三种情况下，分别计算其均值和标准差。

表 2：抽样量改变均值和方差比较

样本量	身高均值	身高 S	体重均值	体重 S
1150	166.68	8.19	57.51	9.43
256	167.60	8.09	60.21	11.64
36	167.5	8.00	57.50	8.41
11	164.36	7.92	53.60	8.44

所有调查数据符合正态分布的前提条件，可采用假设检验，比较两种抽样量的均值和标准差有无显著性差异。从而判断抽样量较小的取值是否可取。若是可取，对将来的类似调查可以节省大量经费。

1.3 假设检验

分别假设检验样本 n_1 、 n_2 和 n_3 与假设总体 $n_0=1150$ 的均值做假设检验， n_1 用如下公式

$H_0: \mu_1=\mu_0$, $H_1: \mu_1\neq\mu_0$ 此时对应 n_1 的假设检验， $\delta=0$ ，若 $t>t_{\alpha/2}(n-1)$ ，则拒绝原假设。

按照 GB/T8053 的推荐，取 $\alpha=0.05$ ，用公式 $s_w=\sqrt{((n_1-1)s_1^2+(n_2-1)s_2^2)/(n_1+n_2-2)}$ 和判断公式 $t=\frac{((\bar{X}_1-\bar{X}_2)-\delta)/S_w\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}}{}$ ，若其小于等于判定数 k ，则认为可以接受原假设^[7]，两者之间无显著性差异。依次对 n_2 和 n_3 重复上述步骤。

查表可得判定数 k 分别是 $t_{0.025}(36)=2.0281$ 和 $t_{0.025}(11)=2.2010$ ， $n=256$ 时，由于表中无法提供自由度大于 45 的 t 值来进行比较，数值只能提供参考。

根据身高的数据由计算得： $t_1=1.6284$ ，由体重的均值代入上式， $t_1=3.9577$ 此时可以参考表中判断数，但无法直接比较。

下表 3 中，对应 $n=36$ 和 $n=11$ 的判断数均小于判断数，接受原假设。下表 3 是假设检验的判断数

表 3：均值的假设检验判断数

k	n=256	n=36	n=11
身高均值检验	1.6284	0.1898	0.936
体重均值检验	3.9577	0.002	1.369

由此可见，在均值的比较上，样本量 $n_2=36$ 和 $n_3=11$ 对应的抽样调查结果都是在可接受范围内，与对照的 $n=1150$ 对应的均值没有显著性差异。根据判断数的显示和对应规则，可以发现， $n_2=36$ 的抽样量对应的结果与对样本的差异性是最小的。

用卡方分布进行方差的比较，用公式： $(n-1)s^2/\sigma_0^2\leq\chi_{1-\alpha/2}^2(n-1)$ 或 $(n-1)s^2/\sigma_0^2\geq\chi_{\alpha/2}^2(n-1)$ 作为拒绝域的判断。分别用 $n_2=36$ 和 11 代入，计算结果如下表 4：

表 4: 方差的假设检验判断数

k	n ₁ =256	n=36	n=11
身高方差检验	/	33.40	9.35
体重方差检验	/	27.84	8.01

查表可得 $n=36$, $\alpha=0.05$ 时的拒绝域为 ≤ 22.106 或 ≥ 54.437 ; 查表可得 $n=11$, $\alpha=0.05$ 时的拒绝域为 ≤ 3.816 或 ≥ 21.920 。可以比照出, 样本量 $n_2=36$ 和 $n_3=11$ 对应调查结果与对照样本的标准差无显著性差异。

由于 χ 分布的表中, 最多提供的就是 $n=45$ 的情况, 所以根据常用表格无法查出拒绝域, 但可以推断出仍在可接受范围内。

两者比较下来都是 $n_2=36$, 和 $n=11$ 的抽样量其方差的假设检验都在可接受范围内。由此可见抽样量的大小对结果的影响并不大, 只要抽样方法是合适的。

在对比了两个抽样量的均值和标准差后, 我们仔细比较了结果的差异大小, 以及差异的显著性。如果直接根据计算结果来判断的话, 这里显示出来反而是抽样量较小的情况下, 其结果与我们的设定的比较均值更为接近。当然, 此时抽样量稍大的情况下, 其均值与设定的比较值也是比较接近的。就是说, 在这个研究对象中, 我们看不出抽样量是 36 与抽样量是 11 的情况有多大差别。特别地, 即使与抽样量为 256 的情况相比, 我们也认为其差别不显著。

同样的, 各抽样量下标准差的差异情况, 我们也通过了假设检验。根据假设检验的计算结果, 也是样本量小的情况下, 其标准差与我们的设定比较值最为接近。当然, 抽样量稍大的情况也是能通过检验的。在这个研究对象下, 我们也没有发现抽样量 36 与 11 之间有多大的差异。

2 结论

根据数理统计规律, 样本量增加呈直线递增的情况下(样本量增加一倍,成本也增加一倍), 而抽样误差只是样本量相对增长速度的平方根递减^[8]。若样本量过大,调查单位增多,不仅增加人力、财力和物力的耗费,增加调查费用,而且还影响到抽样调查的时效性,从而不能充分发挥抽样调查的优越性^[9]。因此,为节省调查费用,体现出抽样的优越性,在确定样本容量时,应在满足抽样对估计数据的精确度的前提下,尽量取小一些的样本量^[10]。

样本量如何选择能保证结果在误差范围内,又能体现抽样的经济性?我们发现,当采用经验样本量的估计公式($n=1.96\sigma/d^2$)来估计样本量时,可以采用相应的国家标准的推荐方法来辅助确定。大多数时候,这两种方法估计出来的样本量没有本质的差异。但当误差设定 d 比较小的时候,两种方法的样本量估计值会相差比较大。抽样量的估计公式($n=1.96\sigma/d^2$)本身就不能提供非常精确的调查,所有把 d 设定得过小,是没有科学意义的。

经验证,样本量大的有时反而差异更显著性。所以在样本量的选择上,推荐合适的不用太大的样本量。这个界限值可采用 GB/T8054-2008 的推荐值来参考,一般选择不大于 60 的样本量。

抽样调查的数据量不能盲目求大求多,这样就失去了抽样的意义,弱化了抽样设计的科学性。同时,不建议盲目追求较小的误差,公式中可以实现误差较小情况下,抽样量的推荐,但实际调查中,过分追求较小的误差 d ,其实是没有实际意义的。如果误差在合理范围内,即误差与标准差几乎相等时,此时抽样量和误差都能控制在适宜的范围,且能做到既经济又合理。这为将来我们对大学生群体的这类调查的抽样量直接提供了参考依据。能大大减少抽样量,减少经费预算和人力投入。

3 参考文献

- [1] 时涛.抽样调查中样本量的科学确定[B].泰山医学院学报,2010-07
- [2] 汤伟伟.走向营销第一步[M].北京:首都师范大学出版社,2009.
- [3] 俞礼军.中小城市起讫点(OD)调查最有抽样量模型.长安大学学报(自然科学版),2008-01

- [4] 上海质量管理科学研究院.顾客满意度测评[M]. 上海科学技术出版社, 2001
- [5] GB/T 8051-2008 计数序贯抽样检验方案
- [6] GB/T 8054-2008 计量标准型一次抽样检验程序及表
- [7] 盛骤.概率论与数理统计, 浙江大学,
- [8] 国家质检总局质量管理司,清华大学中国企业研究中心编著,赵平主编.中国顾客满意指数指南[M]. 中国标准出版社,

2003

- [9] 袁建文, 李科研.关于样本量计算方法的比较研究.理论新探, 2013-01
- [10] 崔颖安.社会化媒体大数据多阶段整群抽样方法.软件学报, 2014,25 (4) : 781-796

【作者简介】

胡小丽, 南京信息职业技术学院副教授, 多年来一直从事关于计量和统计方面的研究。吴文伶, 南京信息职业技术学院副教授。