

敏感问题调查模型浅析

□ 田茂茜

摘要：本文针对敏感性问题造成的无回应给出了要调查问题所占比例的点估计，证明了该估计是无偏估计，几乎处处收敛，依概率收敛，而且从理论上证明了当黑球比例与白球比例的差的绝对值越大时，得到的估计值越接近实值。

关键词：敏感性问题；抽样调查

无论是进行社会问题研究还是进行市场需求的统计分析研究，运用的最为普遍的是抽样调查。显然，抽样调查虽然是非全面调查，但它的目的却在于取得反映总体情况的信息资料，因而，也可起到全面调查的作用。抽样调查可以分为两类，即概率抽样和非概率抽样。

现在被广泛应用的抽样调查是概率抽样。因此，现代的抽样调查是指概率抽样，其定义为：抽样调查，又称抽样推断，是一种重要的、科学的非全面调查方法。它根据调查的目的和任务要求，按照随机原则，从若干单位组成的事物总体中，抽取部分样本单位来进行调查、观察，用所得到的调查标志的数据来推断总体。抽样调查按抽样的组织形式划分，有以下几种主要方法：（1）简单随机抽样（也叫纯随机抽样，SPS 抽样）。（2）等距抽样（也叫机械抽样或系统抽样，SYS 抽样）。（3）类型抽样（也叫分层抽样，STR 抽样）。（4）整群抽样（又称集团抽样）。（5）多阶抽样（又称多级抽样）。（6）二重抽样（又称两相抽样）。（7）比率抽样（PPS 抽样）。

抽样调查中存在的主要问题：1. 确定样本框难度大。2. 被调查者不配合。3. 调查者素质较低，能力差。4. 问卷调查表设计不合理。针对上述提到的问题，我们应该认真研究合适的调查方法和技巧：（1）健全调查网络，建立稳固的调查队伍提升调查员的业务素质，这是减少调查误差的关键。（2）宣传统计法律，打消被调查者的心理顾虑，这是取得准确调查数据的基础。（3）采用迂回访谈，经常换位思考，消除防范心理，从而获得准确数据。（4）科学判断评估。抽样调查除了本身存在的抽样误差外，还存在调查误差。因此对调查数据质量进行科学评估显得尤为重要。

大部分的抽样调查，都会遇到随机抽样误差以外的误差，这些误差会导致产生偏差，使得置信叙述没有意义。产生误差的原因主要有：涵盖不全，即在选样本过程中，如果总体当中的有些部分，根本未被纳入选择范围，这时就发生了涵盖不全的问题，即使从该总体中随机抽样，所得结果还是有偏的；无回应，即无法得到已经被选入样本的个体的资料，最常发生无回应的原因，是联络不上受访对象或者受访对象拒绝回答。一项抽样调查中所宣称的误差界限只包括随机抽样误差。涵盖不全、无回应以及其他实际困难也会造成较大的偏差，但是误差界限并没有包含这些项目在内。从而，我们应尽可能减少这方面造成的偏差。好的技巧都有减少误差的作用。

当调查的问题比较敏感时，被调查者一般不愿意回答，或者即使同意回答但是却做出虚假的答复，这都会造成无回应的发生。例如涉及“你是否赌博”，“曾经是否作弊”，运动员“是否服用兴奋剂”的问题，有过如此经历的人是不愿意承认的。

我们以调查体育运动员服用兴奋剂所占的比例为例，为了得到实际的值，调查人员让运动员在无人场所内，从装有黑球和

白球的袋子中任意取一球，观察其颜色后放回，并承诺若取得黑球就讲真话，取得白球就讲假话，被调查者只需要在匿名的调查表中选“是”或者“否”即可。其中，袋中黑球和白球的比例分别是 p_0 ， q_0 （ $p_0 + q_0 = 1$ ）。

下面的问题是如何得到 p 值的估计 p ， p 是否 p 的无偏估计，为了得到更精确的估计，我们的抽样调查方法应该如何改进。

一、 p 值的估计 p

对于任意一名运动员，设事件 $A_1 =$ “回答曾服用兴奋剂”，事件 $A_2 =$ “实际服用兴奋剂”，利用全概率公式得到

$$\begin{aligned} p_1 &= P(A_1) = P(A_1 | A_2) P(A_2) + P(A_1 | \bar{A}_2) P(\bar{A}_2) \\ &= p_0 P(A_2) + q_0 P(\bar{A}_2) \\ &= p_0 P(A_2) + q_0 (1 - P(A_2)) \\ &= q_0 + (p_0 - q_0) P(A_2), \end{aligned}$$

于是，当 $p_0 \neq q_0$ 时，我们有

$$p = P(A_2) = \frac{P(A_1) - q_0}{p_0 - q_0} = \frac{p_1 - q_0}{p_0 - q_0}.$$

根据概率的统计定义：若是调查了 n 名运动员，其中有 m 名运动员回答“是”，则我们用频率来近似概率，便有 $p_1 = m/n$ 。从

而得到 p 的估计值 $p = \frac{p_1 - q_0}{p_0 - q_0}$ 。

例：在运动员服用兴奋剂比例的调查中，设袋中黑球与白球的比例分别是 $p_0 = 3/10$ ， $q_0 = 7/10$ ，运动员中回答“是”的比例是 $p_1 = 19/30$ ，从而我们得到 p 的估计值

$$p = \frac{19/30 - 7/10}{3/10 - 7/10} = \frac{1}{6}.$$

二、 p 是 p 的无偏估计

命题 1：当 $p_0 \neq q_0$ 时， p 的估计 p 是 p 的无偏估计。

分析：要得到 p 是 p 的无偏估计，只需证明 $E_p = p$ ，又 $E_p = \frac{E p_1 - q_0}{p_0 - q_0}$ ，于是只需要求出 $E p_1$ 即可。

证明：设 X_1, X_2, \dots, X_n 是独立同分布随机变量序列，其中

$$X_i = \begin{cases} 1, & \text{第 } i \text{ 名运动员回答是} \\ 0, & \text{第 } i \text{ 名运动员回答否} \end{cases}, i = 1, 2, 3, \dots, n,$$

则有

$$p_1 = \frac{X_1 + X_2 + \dots + X_n}{n},$$

且 $E X_i = P(X_i = 1)$

$$\begin{aligned} &= P(X_i = 1 | \text{取到黑球}) P(\text{取到黑球}) + P(X_i = 1 | \text{取到白球}) P(\text{取到白球}) \\ &= p p_0 + (1 - p) q_0, \end{aligned}$$

从而有

$$E p_1 = E \left(\frac{X_1 + X_2 + \dots + X_n}{n} \right) = E X_i = p p_0 + (1 - p) q_0,$$

所以

$$E p = \frac{p p_0 + (1 - p) q_0 - q_0}{p_0 - q_0} = p.$$

我国民营上市公司会计信息披露问题研究

□ 张天阳 易剑雯

一、中国民营上市公司会计信息披露的现状分析

上海证券交易所的研究发现：2003～2004年，我国民营上市公司会计信息披露的违规比例明显高于上市公司总体的违规比例，并且呈上升趋势。2003年我国民营上市公司的违规比例为12.18%，明显高于上市公司总体的违规比例7.40%。2004年，民营上市公司会计信息披露的违规比例上升到16.51%，而上市公司总体违规比例为8.51%。另外，我国民营上市公司会计信息披露重复违规的比例也明显高于上市公司总体的重复违规比例。在此期间，民营上市公司的会计信息披露重复违规比例为26.14%，而上市公司总体的重复违规比例为21.63%，两者相差近五个百分

点。我国民营上市公司会计信息披露违规类型主要表现在以下几个方面。

1. 会计信息披露虚假或严重误导性陈述

根据上海证券交易所的研究，2002年～2004年，我国共有11家民营上市公司因会计信息披露虚假性陈述而受到中国证监会的公开处罚、公开批评或证券交易所的公开谴责。

例如：上海腾达智能股份有限公司在2002年12月1日到2003年4月28日期间在董事长的指使下，以不同的方式伪造了大量会计及业务记录。伪造的销货记录、售货记录的金额占腾达2002年全年交易额的40%，致使腾达智能2002年的纯利润被夸大了4倍，虚增了3470万港元。董事

长许培新也因串谋使用虚假文书罪名成立而被判入狱4年（王连娟、龙成凤、孙琦，2006）。

2. 对不当关联交易及对外担保、控股股东的资金占用等重大会计事项披露不足

例如：根据上海证券交易所研究中心的研究报告，湖南国光瓷业集团股份有限公司会计信息披露存在以下违规事实：1. 截至2004年8月31日，湖南国光瓷业集团股份有限公司对其控股股东上海鸿仪投资发展有限公司及其关联公司占用的资金22321.15万元，没有履行相应的审议程序，也没有履行临时公告的会计信息披露义务。2. 截至2004年8月31日，湖南国光瓷业及其控股子公司对外提供的担保金额共51218.31万元没有履行临时公告的

命题2：当 $p_0 \neq q_0$ 时，对于 p 值的估计 p ，有 $p \rightarrow p$, a. s.。

证明：由强大数定律 [1]，我们有 $p_1 \rightarrow EX_1$ a. s.，即存在样本空间 Ω 中的点集 Ω_0 ，满足 $P(\Omega_0) = 0$ ，且 $\lim_{n \rightarrow \infty} p_1 = EX_1$ ， $\omega \in \Omega \setminus \Omega_0$ ，从而有

$$\lim_{n \rightarrow \infty} \frac{p_1 - q_0}{p_0 - q_0} = \frac{EX_1 - q_0}{p_0 - q_0} = p, \omega \in \Omega \setminus \Omega_0 \text{ 或 } \lim_{n \rightarrow \infty} p = p, \omega \in \Omega \setminus \Omega_0,$$

即 $p \rightarrow p$, a. s.。

几乎必然收敛和依概率收敛这两种收敛性分别对应强大数定律和弱大数定律，强大数定律比弱大数定律要强，从而几乎必然收敛可以推得依概率收敛 [2]，我们得到下面的结论：

推论：当 $p_0 \neq q_0$ 时，对于 p 值的估计 p ，有 $p \xrightarrow{P} p_0$ 。

三、如何设计才能减少误差

命题3：当 $p_0 \neq q_0$ ，且 $|p_0 - q_0|$ 增大时， p 的方差 Dp 减小。

证明： $Dp = D\left(\frac{p_1 - q_0}{p_0 - q_0}\right) = \frac{Dp_1}{(p_0 - q_0)^2}$ ，由命题1的证明过程知

$$Dp_1 = \frac{DX_1}{n} = \frac{EX_1 - (EX_1)^2}{n} = \frac{pp_0 + (1-p)q_0 - (pp_0 + (1-p)q_0)^2}{n},$$

从而有

$$Dp = \frac{Dp_1}{(p_0 - q_0)^2} = \frac{pp_0 + (1-p)q_0 - [pp_0 + (1-p)q_0]^2}{n(p_0 - q_0)^2} = \frac{p(p_0 - q_0) + q_0 - p^2(p_0 - q_0)^2 - 2pq_0(p_0 - q_0) - q_0^2}{n(p_0 - q_0)^2} = \frac{p(p_0 - q_0) - p^2(p_0 - q_0)^2 + p_0q_0 - 2pq_0(p_0 - q_0)}{n(p_0 - q_0)^2}$$

$$= \frac{p(p_0 - q_0)^2 - p^2(p_0 - q_0)^2 + p_0q_0}{n(p_0 - q_0)^2} = \frac{1}{n} \left[(p - p^2) + \frac{p_0q_0}{(p_0 - q_0)^2} \right]$$

$$\text{又 } \frac{p_0q_0}{(p_0 - q_0)^2} = \frac{(p_0 + q_0)^2 - (p_0 - q_0)^2}{4(p_0 - q_0)^2} = \frac{1}{4(p_0 - q_0)^2} - \frac{1}{4}, \text{ 代入上式得}$$

$$Dp = \frac{1}{n} \left[(p - p^2) + \frac{1}{4(p_0 - q_0)^2} - \frac{1}{4} \right].$$

由此可见，当 $|p_0 - q_0|$ 增大时， p 的方差 Dp 减小。

当 $|p_0 - q_0|$ 增大时， p 会更接近 p ，但是往往也不会为被调查者所接受，无回应还是会发生。遇到这样的情况，我们应该怎样处理这些问题呢？

首先，用其他人取代不回应的人。因为城市里的不回应率很高，如果用不回应住户附近的其他住户来取代，可以减低偏差。其次，数据搜集完成之后，应该用统计方法给回应加权，以纠正误差来源。如果城市里太多住户没回应，就给城市里有回应的住户加权。如果样本里太多女性，就给男士们加权。加权的确可以修正偏差，但是也会增加变异性。这就需要统计学家们继续研究更多地调查方法。

参考文献：

- [1] 茆诗松，程依明，濮晓龙. 概率论与数理统计教程 [M]. 北京：高等教育出版社，2004.
- [2] 李贤平，沈崇圣，陈子毅. 概率论与数理统计 [M]. 上海：复旦大学出版社，2003. 251.
- [3] 戴维. S. 穆尔著，郑维厚，译. 统计学的世界（第五版）. 北京：中信出版社，2003.