

参数区间估计和样本容量确定在市场调查中的应用

周莉莉, 朱美玲

(新疆农业大学经贸学院, 乌鲁木齐 830052)

摘要 简述了参数区间估计的概念、方法以及样本容量的确定方法。在此基础上, 给出市场调查中比率的区间估计、样本容量确定和均值的区间估计实例。

关键词 区间估计 样本容量 市场调查

中图分类号 O212

文献标识码 A

市场调查能够使我们及时掌握市场信息和市场动向, 为企业决策和国民经济宏观管理提供所需的信息。实际上由于调查经费和调查时间等条件的限制, 市场调查往往不可能采用全面调查的形式, 而只能对调查总体中的一部分单位进行调查, 再由部分推断总体, 这就是抽样调查方法。抽样调查所获取的是样本的数据, 而我们进行市场调查最终是要获得有关总体的信息, 所以必须用样本的数据来推断总体的指标, 而且, 为了获得具有一定精度和可靠度的总体指标, 我们必须确定必要的样本容量, 保证估计的精度和可靠性。

1 参数的区间估计和样本容量的确定方法

1.1 区间估计

设 $x_1, x_2, \dots, x_n$ 是来自 $f(x, \theta)$ 的样本, 对给定的 $\alpha (0 < \alpha < 1)$, 若有统计量 $\theta_1(x_1, x_2, \dots, x_n)$ 及 $\theta_2(x_1, x_2, \dots, x_n)$ 使 $P\{\theta_1(x_1, x_2, \dots, x_n) < \theta < \theta_2(x_1, x_2, \dots, x_n)\} = 1 - \alpha$, 则称 $1 - \alpha$ 是置信度或置信概率 (Confidence probability); $(\theta_1(x_1, x_2, \dots, x_n), \theta_2(x_1, x_2, \dots, x_n))$ 是置信度为 $1 - \alpha$ 的 θ 置信区间 (Confidence interval); α 称为显著性水平 (Significance level)。

置信区间的长度表达了区间估计的精确性, 置信概率表达了区间估计的可靠性。在进行区间估计时, 必须同时考虑置信概率和置信区间两个方面, 一般地, 置信概率定得越大, 则置信区间相应也越大, 所以, 可靠性与精确性要结合具体问题、具体要求全面考虑。

1.2 总体均值的区间估计

1.2.1 总体方差 σ^2 已知时总体均值 μ 的区间估计

当 $X \sim N(\mu, \sigma^2)$ 时, 可以证明抽自该总体的简单随机样本 $x_1, x_2, \dots, x_n$ 的样本均值 $\bar{x}$ 服从数学期望为 μ , 方差为 σ^2/n 的正态分布, 即

$$\bar{x} \sim N(\mu, \sigma^2/n)$$

当总体方差 σ^2 已知时, 建立置信区间所用的统计量是 Z 统计量:

$$Z = (\bar{x} - \mu) / (\sigma / \sqrt{n}) \sim N(0, 1)$$

收稿日期: 2001 - 09 - 28

根据区间估计的定义,可以构造均值 μ 的置信区间。

对于给定的显著性水平 α ,可以令

$$P\{-Z_{\alpha/2} < Z < Z_{\alpha/2}\} = 1 - \alpha$$

$$\text{即有 } P\left\{-Z_{\alpha/2} < (\bar{x} - \mu)/(\sigma/\sqrt{n}) < Z_{\alpha/2}\right\} = 1 - \alpha$$

$$\text{从而有 } P\left\{\bar{x} - Z_{\alpha/2}(\sigma/\sqrt{n}) < \mu < \bar{x} + Z_{\alpha/2}(\sigma/\sqrt{n})\right\} = 1 - \alpha$$

即在给定显著性水平 α 下,总体均值 μ 在 $1 - \alpha$ 的置信水平下的置信区间为:

$$[\bar{x} - Z_{\alpha/2}(\sigma/\sqrt{n}), \bar{x} + Z_{\alpha/2}(\sigma/\sqrt{n})]$$

当总体为非正态总体时,可以证明,当样本容量 n 足够大时,样本均值 $\bar{x}$ 近似地服从数学期望为 μ ,方差为 σ^2/n 的正态分布,经验上规定 $n > 30$ 就是样本容量足够大了。

1.2.2 总体方差 σ^2 未知时总体均值 μ 的区间估计

当总体服从正态分布,但总体方差 σ^2 未知时,要用样本方差 S_{n-1}^2 代替 σ^2 来建立置信区间。这时,新的统计量不服从标准正态分布,而是服从自由度为 $n - 1$ 的 t 分布,记为:

$$t = (\bar{x} - \mu)/(S_{n-1}/\sqrt{n}) \sim t(n - 1)$$

此时总体均值的置信区间为:

$$[\bar{x} - t_{\alpha/2}(S_{n-1}/\sqrt{n}), \bar{x} + t_{\alpha/2}(S_{n-1}/\sqrt{n})]$$

当总体为非正态总体且 σ^2 未知时,只要样本足够大,一般当 $n \geq 30$ 时,仍可用上式来近似地建立总体均值 μ 的置信区间。

1.2.3 总体比例的区间估计

我们称总体中具有某种特征的单位占总体全部单位的比例为总体比例,记为 P ;称样本具有该种特征的单位占样本全部单位的比例为样本比例,记为 p 。

可以证明,在大样本下,若 $np > 5$, $n(1 - p) > 5$,则可以把二项分布问题变换为正态分布问题近似地去求解。因而有:

$$p \sim N[P, P(1 - P)/n]$$

即样本比例 p 服从期望值为 P ,方差为 $P(1 - P)/n$ 的正态分布。因而,可以用 Z 统计量来构造总体比例 P 的置信区间,即

$$Z = (p - P)/\sqrt{P(1 - P)/n} \sim N(0, 1)$$

在估计 P 时,由于 P 未知,则用样本比例 p 代替 P 计算估计量的标准误差,在 $1 - \alpha$ 的置信水平下,总体比例 P 的置信区间为:

$$(p - Z_{\alpha/2}\sqrt{p(1 - p)/n}, p + Z_{\alpha/2}\sqrt{p(1 - p)/n})$$

1.3 样本容量的确定

在以上的叙述中,我们假定样本容量(Sample size) n 是已知的,但在实际问题中,需要自己设计调查方案及决定样本容量的大小。样本容量 n 选得越大,估计精度越高,但费用会相应增加;如果 n 选得过小,估计误差增大,可靠度降低。样本容量的大小主要依据估计需要的精度与可靠度来选择。

1.3.1 估计总体均值时,样本容量的确定

在总体均值的区间估计里,置信区间为:

$$[\bar{x} - Z_{\alpha/2}(\sigma/\sqrt{n}), \bar{x} + Z_{\alpha/2}(\sigma/\sqrt{n})]$$

从估计量 $\bar{x}$ 到 $\bar{x} - Z_{\alpha/2}(\sigma/\sqrt{n})$ 和从 $\bar{x}$ 到 $\bar{x} + Z_{\alpha/2}(\sigma/\sqrt{n})$ 的距离表示在一定置信水

平 $1 - \alpha$ 下,用样本均值估计总体均值时所允许的绝对误差,我们用 Δ 表示,称 Δ 为允许误差 (Permissible error)。显然,若以 $\bar{x}$ 的取值为中心,则允许误差 Δ 可以表示为:

$$\Delta = Z_{\alpha/2}(\sigma/\sqrt{n})$$

对于一个具体问题,总体标准差 σ 是一个确定数值,给出了我们所希望的“可靠程度”(即置信水平 $1 - \alpha$),可靠性系数 $Z_{\alpha/2}$ 也就唯一地被确定了。当给出了允许误差后,必要的样本容量就可给出:

$$n = Z_{\alpha/2}^2 \sigma^2 / \Delta^2$$

由此可看出:样本容量 n 与总体方差 σ^2 成正比,与允许误差 Δ^2 成反比,与可靠性系数 $Z_{\alpha/2}^2$ 成正比。

当总体方差 σ^2 未知时,要用样本方差 S_{n-1}^2 来估计总体的方差 σ^2 ,此时

$$n = t_{\alpha/2}^2 S_{n-1}^2 / \Delta^2$$

1.3.2 估计总体比例时,样本容量的确定

在估计总体比例时,当总体的比例 P 已知时,必要的样本容量为:

$$n = Z_{\alpha/2}^2 P(1 - P) / \Delta^2$$

当总体的比例 P 未知,则用经验比例 p 代替 P 计算总体比例的方差。如果没有可以利用的 p 估计值,对于服从二项分布的随机变量,当 $p = 0.5$ 时,其方差达到最大值,可以用 $p = 0.5$ 计算,这样得出的必要样本容量虽然可能比实际需要的容量大一些,但可以充分保证有足够高的置信水平和尽可能小的置信区间。必要的样本容量为:

$$n = Z_{\alpha/2}^2 p(1 - p) / \Delta^2$$

2 应用实例

2.1 实例一

在2001年5月,我们运用抽样调查的方法对中国联通公司新疆分公司在乌鲁木齐市某地区的市场占有率情况进行了调查,并运用参数区间估计的方法对所调查范围的总体情况进行推断。

2.1.1 样本容量的确定

我们欲估计乌鲁木齐市某地区126台寻呼机和130手机的使用率,希望对使用率 P 的估计误差不超过0.05,要求的可靠度为90%,应取的样本容量的估算:

$$\text{即 } \Delta = 0.05, \alpha = 0.10, Z_{\alpha/2} = 1.64, p = 0.50$$

$$n = Z_{\alpha/2}^2 p(1 - p) / \Delta^2 = 269(\text{人})$$

为了以90%的可靠度保证估计误差不超过0.05,应取269个人进行访问调查。考虑到访问无效的情况,我们把被调查者人数增至310人,回收的有效问卷为300份,足以满足精度和可靠度的要求。

2.1.2 置信区间的估计

126寻呼台使用率区间估计

根据样本数据算得126寻呼台的使用率是18.94%,置信度为90%,即 $\alpha = 0.10$,已知 $n = 300$, $p = 0.1894$, $np = 57 > 5$, $n(1 - p) = 243 > 5$,当 $\alpha = 0.10$ 时, $Z_{\alpha/2} = 1.64$,

$$\{p - Z_{\alpha/2} \sqrt{p(1 - p)/n}, p + Z_{\alpha/2} \sqrt{p(1 - p)/n}\} = (0.1523, 0.2265)$$

即在乌鲁木齐市某地区 126 台的使用率在 15.23 % 至 22.65 % 之间,作出这一估计的可靠程度为 90 %。

130 手机使用率区间估计

根据样本数据计算得 130 手机使用率为 28.65 %,同样将置信度定为 90 %, $n = 300$, $p = 0.2865$, $np = 86 > 5$, $n(1 - p) = 214 > 5$,当 $\alpha = 0.10$ 时, $Z_{\alpha/2} = 1.64$,

$$[p - Z_{\alpha/2} \sqrt{p(1 - p)/n}, p + Z_{\alpha/2} \sqrt{p(1 - p)/n}] = (0.2437, 0.3293)$$

即在乌鲁木齐市某地区 130 台的使用率在 24.37 % 至 32.93 % 之间,作出这一估计的可靠程度为 90 %。

2.2 实例二

2001 年 6 月,我们运用抽样调查的方法对新疆农业大学 14 个学院、系、部的三个年级(98 级、99 级、2000 级)的在校学生进行了网络知识了解程度的调查,调查结束后,运用参数区间估计的方法对这三个年级的全体学生对网络知识的了解程度进行了推断。

2.2.1 确定样本容量

确定区间估计的可靠度为 90 %,对网络知识是否了解的比率 P 的估计误差不超过 3 %,即 $\alpha = 0.10$, $Z_{\alpha/2} = 1.64$, $\Delta = 0.03$, $p = 0.50$

$$n = Z_{\alpha/2}^2 p(1 - p) / \Delta^2 = 747 (\text{人})$$

应抽取 747 个学生进行调查,考虑到会出现无效问卷的情况,我们将被调查者人数增至 780 人,最终回收的有效问卷为 710 份,比所要求的样本容量略少。

2.2.2 学生对网络知识的了解程度的区间估计

问卷中有关网络知识的问题共 7 道,调查结果见表 1。根据答对问题的数目,我们对学生对网络知识的了解程度分成 4 个层次。

表 1 学生对网络知识的了解程度问卷调查表

答对问题数	人数(人)	比率(%)	了解程度
0	48	6.76	完全不了解
1	46	6.48	
2	49	6.90	
3	92	12.96	了解一点
4	108	15.21	
5	158	22.25	
6	125	17.61	比较了解
7	84	11.83	
合计	710	100	—

根据表 1 的调查结果,计算的样本学生平均答对题数 $\bar{x} = 4.197$ 道,标准差 $S_{n-1} = 2.007$ 道,可靠度定为 90 %,则学生总体平均答对题数的区间为: $[\bar{x} - t_{\alpha/2}(S_{n-1}/\sqrt{n}), \bar{x} + t_{\alpha/2}(S_{n-1}/\sqrt{n})] = (4.073, 4.321)$

因此,学生总体对网络知识的了解程度仅为比较了解,还需要加强对学生网络知识的传

授,作出这一估计的可靠度为 90 %。

3 结 论

从以上理论方法及应用实例可以看出,运用样本容量的确定方法和参数区间估计的方法可以有效地控制用样本资料推断总体数据的可靠度和精确度。在市场调查中广泛地运用抽样调查的方法,并利用样本容量确定和区间估计的方法既能够节省调查费用和缩短调查时间,又能够获得具有一定可靠程度和精度的总体数据,实际运用效果非常好。

参考文献

- 1 贾俊平,何晓群,金勇进. 统计学. 北京:中国人民大学出版社,2000.

Interval Estimator and Sample Size Defining in the Application of Marketing Survey

Zhou Lili , Zhu Meiling

(College of Economics and Trade ,Xinjiang Agricultural University , Urumqi ,830052)

Abstract This article simplified the concept and the method of interval estimator and the method of defining the sample size. On the basis of the above-mentioned content , two examples are given for applying these methods.

Key words interval estimator ; sample size ; marketing survey.