

## 数量特征敏感问题加法 RRT 模型下分层三阶段抽样的统计方法及应用

陈科锦,高 歌\*,范玉波,杨少春,杜俏俏

(苏州大学医学部公共卫生学院流行病与卫生统计学系,江苏 苏州 215123)

**[摘要]** **目的:**为实际需要的数量特征敏感问题复杂抽样提供可靠有效的调查方法及其统计公式;为制订艾滋病的预防控制措施提供科学依据。**方法:**采用数理统计学理论方法推导统计公式;对北京市男男性行为(MSM)人群进行实例调查;通过蒙特卡洛模拟,采用 SAS 编程模拟调查分析 100 个样本,对本研究的调查方法及其统计公式作信度与效度评价。**结果:**推导出数量特征敏感问题加法模型下分层三阶段抽样总体均数估计量及其方差的统计公式;北京市 MSM 人群首次男男性行为的年龄、月不同性伴人数、月性行为次数的均数分别为 21.96 岁、2.80 人、4.85 次,其标准误分别为 0.127 岁、0.096 人、0.559 次;100 个总体均数的 95% 可信区间均包含模拟总体均数。**结论:**MSM 人群具有艾滋病的高危性行为,应加强其预防控制。本文研究的调查方法及其统计公式具有良好的信度与效度,可用于今后大规模的敏感问题抽样调查。

**[关键词]** 敏感问题的随机应答技术;分层三阶段抽样;男男性行为;蒙特卡洛方法;效度与信度

**[中图分类号]** C81

**[文献标志码]** A

**[文章编号]** 1007-4368(2014)08-1115-05

**doi:**10.7655/NYDXBNS20140824

## Additive RRT model for inquiring quantitative sensitive questions under stratified three-stage sampling and its application

Chen Kejin, Gao Ge\*, Fan Yubo, Yang Shaochun, Du Qiaoqiao

(Public Health School, Soochow University, Suzhou 215123, China)

**[Abstract]** **Objective:**To explore reliable and efficient sampling methods and corresponding formulae for quantitative sensitive question survey on three-stage random sampling. To provide scientific data for the prevention and control of high risk AIDS population. **Methods:**We derived statistical formulae from mathematical statistics theory;investigated MSMs in Beijing for instance; used SAS programming to simulate investigation and analysis of 100 samples through Monte Carlo simulation and accessed validity and reliability of methods and formulae studied in this article. **Results:**We derived statistical formulae of the overall mean estimator and its variance on quantitative sensitive questions under stratified three-stage sampling. The means of ages of the first occurrence of MSM, the monthly number of different sexual partners and the monthly number of sex were 21.96 years old, 2.80 people, 4.85 times, respectively, among MSM in Beijing, and their standard errors were 0.127 years old, 0.096 people, 0.559 times, respectively. All of the population mean of 95% confidence interval obtained from 100 analog samples contains population mean. **Conclusion:**For high-risk sexual behavior among MSM with AIDS, its prevention and control should be strengthened. Monte Carlo computer simulation results show that stratified three-stage sampling method and corresponding formulae are feasible and could be applied in the sensitive issue of large-scale sampling in the future.

**[Key words]** Randomized response technique on sensitive questions; stratified three-stage sampling; MSM; Monte Carlo simulation; validity and reliability

[Acta Univ Med Nanjing, 2014, 34(08):1115-1119]

抽样调查作为调查研究方法,近年来在医学科研和卫生工作、政府各部门、企业与学术界都有了极

大的发展<sup>[1-3]</sup>。各式各样的高度私人机密性或大多数人认为不便在公开场合表态及陈述的问题,即敏感性问题,会在抽样调查研究中经常遇到<sup>[4]</sup>。敏感问题涉及到药物与酒精滥用、性传播疾病、家庭暴力、偷税漏税等社会现象的方方面面<sup>[5]</sup>。敏感性问题按总

**[基金项目]** 国家自然科学基金项目(81273188)

\*通信作者(Corresponding author), E-mail: gaoge@suda.edu.cn

体的特征可分为两类——属性特征敏感性和数量特征敏感性问题<sup>[4]</sup>。对于敏感性问题,若采用传统的直接问答方式,被调查者常常会故意误答,调查者将难以控制样本信息,导致调查结果产生信息偏倚<sup>[6]</sup>。随机应答技术(randomized response technique, RRT)是指在调查过程中使用特定的随机化装置,使被调查者以一个预定的基础概率  $P$  从 2 个或 2 个以上的问题中选择 1 个问题进行回答,或回答非敏感问题的数值,除被调查者以外的所有人(包括调查者)均不知道被调查者的回答是针对哪一个问题或敏感问题的数值,以便保护被调查者的隐私,最后根据概率论的知识计算出敏感问题特征在人群中的真实分布的一种调查方法<sup>[7]</sup>。

本文基于数量特征敏感问题加法 RRT 模型的分层三阶段抽样,从数学上推导出敏感问题总体均数的估计量及其方差的计算公式;并采用本研究的调查方法及统计公式,对北京市男男性行为者(men who have sex with men, MSM) 这一艾滋病高危人群的数量特征敏感问题进行调查分析;根据实际调查的统计量数值作为模拟总体参数,采用 SAS 编程建立模拟总体并模拟调查分析 100 个样本,平均每个样本包含 2 533 例模拟调查对象,对本研究的调查方法及其统计公式进行信度与效度评价。

### 1 数量特征敏感问题加法模型分层三阶段抽样的统计方法

#### 1.1 加法 RRT 模型

设计一套随机装置:在小布袋中放置大小、重量、质感相同的 10 个小球,10 个小球上分别贴有 0、1、2、...、9 的数字标签。被抽中的每个对象(三级单位)从小布袋中有放回地随机抽取 1 个小球,在没有旁人知晓时将自己敏感问题特征的数值( $X$ )与抽中小球上的数值(随机变量  $Y$ )相加结果  $Z$  填入调查表中。

#### 1.2 抽样方法

假定总体划分成  $L$  个层。第  $h$  层由  $N_{1h}$  个一级单位组成; $h$  层第  $i$  个一级单位由  $N_{2h}$  个二级单位组成, $h$  层平均每个一级单位包含  $\bar{N}_{2h}$  个二级单位; $h$  层第  $i$  个一级单位内第  $j$  个二级单位包含  $N_{3h}$  个三级单位, $h$  层平均每个二级单位包含  $\bar{N}_{3h}$  个三级单位, $h$  层共有  $N_h$  个三级单位,总体中共有  $N$  个三级单位;这里, $h=1,2,\dots,L,i=1,2,\dots,N_{1h},j=1,2,\dots,N_{2h}$ 。第一阶段从  $h$  层随机抽取  $n_{1h}$  个一级单位;第二阶段从  $h$  层第  $i$  个被抽中的一级单位内随机抽取

$n_{2h}$  个二级单位,平均从  $h$  层每个被抽中的一级单位内随机抽取  $\bar{n}_{2h}$  个二级单位;第三阶段从  $h$  层第  $i$  个抽中的一级单位内第  $j$  个抽中的二级单位中随机抽取  $n_{3h}$  个三级单位,平均从  $h$  层每个抽中的二级单位中随机抽取  $\bar{n}_{3h}$  个三级单位;这里  $h=1,2,\dots,L,i=1,2,\dots,n_{1h},j=1,2,\dots,n_{2h}$ 。对每个被抽中的三级单位,采用数量特征敏感问题加法模型进行调查。

#### 1.3 总体均数的估计量及其方差

$y_{ijk}$  记  $h$  层第  $i$  个一级单位内第  $j$  个二级单位中第  $k$  个三级单位的指标值,  $\hat{\mu}_{ijh}$  记  $h$  层第  $i$  个一级单位内第  $j$  个二级单位总体平均值  $\mu_{ijh}$  的估计量,  $\hat{\mu}_{ih}$  记  $h$  层第  $i$  个一级单位总体平均值  $\mu_{ih}$  的估计量,  $\hat{\mu}$  记总体平均数  $\mu$  的估计量,  $V(\hat{\mu})$  记  $\hat{\mu}$  的方差,  $h=1,2,\dots,L,i=1,2,\dots,n_{1h},j=1,2,\dots,n_{2h},k=1,2,\dots,n_{3h}$ 。根据高歌等<sup>[8]</sup>的公式可得:

$$\hat{\mu}_{ih} = \frac{\sum_{j=1}^{n_{2h}} N_{ij3h} \hat{\mu}_{ijh}}{\sum_{j=1}^{n_{2h}} N_{ij3h}}, h=1,2,\dots,L,i=1,2,\dots,n_{1h} \quad (1)$$

$$\hat{\mu}_h = \frac{N_{1h}}{N_h n_{1h}} \sum_{i=1}^{n_{1h}} \frac{N_{2h}}{n_{2h}} \sum_{j=1}^{n_{2h}} \frac{N_{ij3h}}{n_{ij3h}} \sum_{k=1}^{n_{ij3h}} y_{ijk} = \frac{N_{1h}}{N_h n_{1h}} \sum_{i=1}^{n_{1h}} \frac{N_{2h}}{n_{2h}} \sum_{j=1}^{n_{2h}} N_{ij3h} \hat{\mu}_{ijh}, h=1,2,\dots,L \quad (2)$$

$$\hat{\mu} = \sum_{h=1}^L \frac{N_h \hat{\mu}_h}{N} = \sum_{h=1}^L W_h \hat{\mu}_h \quad (W_h = \frac{N_h}{N}) \quad (3)$$

$$V(\hat{\mu}) = \sum_{h=1}^L W_h^2 \left[ \frac{\sigma_{1h}^2}{n_{1h}} \left(1 - \frac{n_{1h}}{N}\right) + \frac{\sigma_{2h}^2}{n_{1h} \bar{n}_{2h}} \left(1 - \frac{\bar{n}_{2h}}{N_{2h}}\right) + \frac{\sigma_{3h}^2}{n_{1h} \bar{n}_{2h} \bar{n}_{3h}} \left(1 - \frac{\bar{n}_{3h}}{N_{3h}}\right) \right] \quad (4)$$

其中:

$$\sigma_{1h}^2 = \frac{1}{N_{1h}-1} \sum_{i=1}^{N_{1h}} (\mu_{ih} - \mu_h)^2$$

其样本估计量为:

$$S_{1h}^2 = \frac{1}{n_{1h}-1} \sum_{i=1}^{n_{1h}} (\hat{\mu}_{ih} - \hat{\mu}_h)^2, h=1,2,\dots,L \quad (5)$$

$$\sigma_{2h}^2 = \frac{1}{N_{2h}} \sum_{i=1}^{N_{2h}} \frac{1}{N_{2h}-1} \sum_{j=1}^{N_{2h}} (\mu_{ijh} - \mu_{ih})^2$$

其样本估计量为:

$$S_{2h}^2 = \frac{1}{n_{2h}} \sum_{i=1}^{n_{2h}} \frac{1}{n_{2h}-1} \sum_{j=1}^{N_{2h}} (\hat{\mu}_{ijh} - \hat{\mu}_{ih})^2, h=1,2,\dots,L \quad (6)$$

$$\sigma_{3h}^2 = \frac{1}{N_{3h}} \sum_{i=1}^{N_{3h}} \frac{1}{N_{3h}-1} \sum_{j=1}^{N_{3h}} \frac{1}{N_{ij3h}-1} \sum_{k=1}^{N_{ij3h}} (y_{ijk} - \mu_{ijh})^2$$

$$= \frac{1}{N_{1h}} \sum_{i=1}^{N_{1h}} \frac{1}{N_{i2h}} \sum_{j=1}^{N_{i2h}} N_{ij3h} V(\hat{\mu}_{ijh})$$

其样本估计量为:

$$\begin{aligned} S_{3h}^2 &= \frac{1}{n_{1h}} \sum_{i=1}^{n_{1h}} \frac{1}{n_{i2h}} \sum_{j=1}^{n_{i2h}} \frac{1}{n_{ij3h}-1} \sum_{k=1}^{n_{ij3h}} (y_{ijhk} - \mu_{ijh})^2 \\ &= \frac{1}{n_{1h}} \sum_{i=1}^{n_{1h}} \frac{1}{n_{i2h}} \sum_{j=1}^{n_{i2h}} n_{ij3h} \cdot v(\hat{\mu}_{ijh}), h=1, 2, \dots, L \end{aligned} \quad (7)$$

#### 1.4 $\hat{\mu}_{ijh}$ 、 $v(\hat{\mu}_{ijh})$ 的估计

$\mu_{ijh}$  用表示  $h$  层第  $i$  个一级单位第  $j$  个二级单位敏感问题特征  $X$  的均数,  $\mu_{ijh}$ 、 $\hat{\mu}_{ijh}$ 、 $S_{ijh}^2$  分别表示  $h$  层第  $i$  个一级单位第  $j$  个二级单位所有回答值的总体均数、样本均数、样本方差,  $\mu_Y$  表示随机化装置中所有随机数的均数, 由均数的基本性质推导可得<sup>[9]</sup>:

$$\mu_{ijh} = \mu_{ijhz} - \mu_Y, h=1, 2, \dots, L; i=1, 2, \dots, n_{1h}; j=1, 2, \dots, n_{i2h} \quad (8)$$

$$\text{则: } \hat{\mu}_{ijh} = \hat{\mu}_{ijhz} - \mu_Y \quad (9)$$

对公式(9)按方差的性质及标准误的计算公式有<sup>[9]</sup>:

$$V(\hat{\mu}_{ijh}) = V(\hat{\mu}_{ijhz})$$

$$\text{所以: } v(\hat{\mu}_{ijh}) = v(\hat{\mu}_{ijhz}) = S_{ijhz}^2 / n_{ij3h} \quad (10)$$

## 2 应用实例与信度效度评价

### 2.1 调查设计

#### 2.1.1 调查对象与调查指标

调查对象为 2010 年 9~10 月出入北京市男同性恋活动场所的 15~49 岁 MSM 人群, 总体数量约 67 750 例。调查指标为首次男男性行为的年龄、男男性行为月人均不同性伴人数、月人均男男性行为次数。

#### 2.1.2 抽样方法

本调查采用分层三阶段抽样, 各层内采用三阶段抽样。将北京市 MSM 人群按年龄分为 15~29 岁 (58.24%) 和 30~49 岁 (41.76%) 两层; 在各层内分别以北京的 16 个区县为一级单位,  $N_1=N_2=16$ ; 以男同性恋者活动场所 (如男同性恋浴室、会所、酒吧、公园等) 为二级单位 (调查点), 以男男性行为者作为三级单位; 两个年龄层内平均每个区县男同性恋人数分别约为  $\bar{N}_{11}=2 466$  人,  $\bar{N}_{22}=1 768$  人。样本含量参考 Wang 等<sup>[10]</sup>给出的公式确定。第一阶段从各年龄层的 16 个区县中分别随机抽取 3 个区县 ( $n_{11}=n_{12}=3$ ); 第二阶段在每个区县随机抽取 5 个调查点 ( $n_{21}=n_{22}=5$ ); 第三阶段从被抽中的 15 个调查点中共随机抽取

2 062 人, 在各层内平均从每个抽中调查点分别抽取了约  $\bar{n}_{31}=80$  人,  $\bar{n}_{32}=57$  人。

#### 2.1.3 RRT 模型

采用数量特征敏感问题随机应答模型进行调查, 以首次男男性行为的年龄调查为例, 设计一套随机装置如 1.1 节所述: 在没有旁人知晓时将自己首次男男性行为的年龄 ( $X$ ) 与抽中小球上的数字 (随机变量  $Y=0, 1, 2, \dots, 8, 9$ ) 相加结果  $Z(Z=X+Y)$  填入调查表中。

### 2.2 调查结果

对均数的计算: 先按公式(9)计算  $\hat{\mu}_{ijh}$ , 再按公式(2)计算  $\hat{\mu}_h$ , 从而按公式(3)计算  $\hat{\mu}$ ; 对  $\hat{\mu}$  的方差的计算: 先按公式(1)计算  $\hat{\mu}_{ih}$ , 再按公式(5)、(6)计算  $S_{1h}^2$ 、 $S_{2h}^2$ , 并按公式(10)计算  $v(\hat{\mu}_{ijh})$ , 从而按公式(7)计算  $S_{3h}^2$ , 最后按公式(4)计算  $V(\hat{\mu})$  的样本估计值。全部计算采用 SAS 编程完成, 计算结果为: 北京市 MSM 人群首次男男性行为的年龄均数为 21.96 岁、标准误为 0.127 岁; 男男性行为月人均不同性伴人数 2.80 人, 标准误为 0.096 人; 月人均男男性行为次数为 4.85 次, 标准误为 0.559 次。

### 2.3 信度效度评价

#### 2.3.1 模拟调查分析步骤

以 2.2 节中实际调查得到的统计量数值作为模拟总体的参数。对于数量特征敏感问题“首次发生男男性行为的年龄”, 用 SAS 编写宏程序 %sample 模拟 100 个分层三阶段抽样下数量特征敏感问题加法 RRT 模型的调查样本, 平均每个样本包含 2 533 例模拟调查对象。并对模拟调查数据根据本文第一部分推导出的统计公式, 计算 100 个样本均数及其估计方差、总体均数的 95% 可信区间。编写 SAS 宏程序代码, 填入相应的宏参数, 并执行编写的宏程序。

#### 2.3.2 模拟调查分析结果

对于数量特征敏感问题“第 1 次发生男男性行为时的年龄”, 模拟总体均数为 21.96 岁 (图 1 中的中横线)。以模拟的样本 (100 个) 序号为横坐标, 以每个样本所得的总体均数 95% 可信区间为纵坐标画图 (图 1)。

根据图 1 可知, 100 个模拟样本得到的 100 个总体均数 95% 可信区间全部包含模拟总体均数, 可认为 100 个样本均数与总体均数 (模拟真值) 的差别无统计学意义, 说明本研究的分层三阶段抽样下数量特征敏感问题加法模型的调查方法及其统计公式

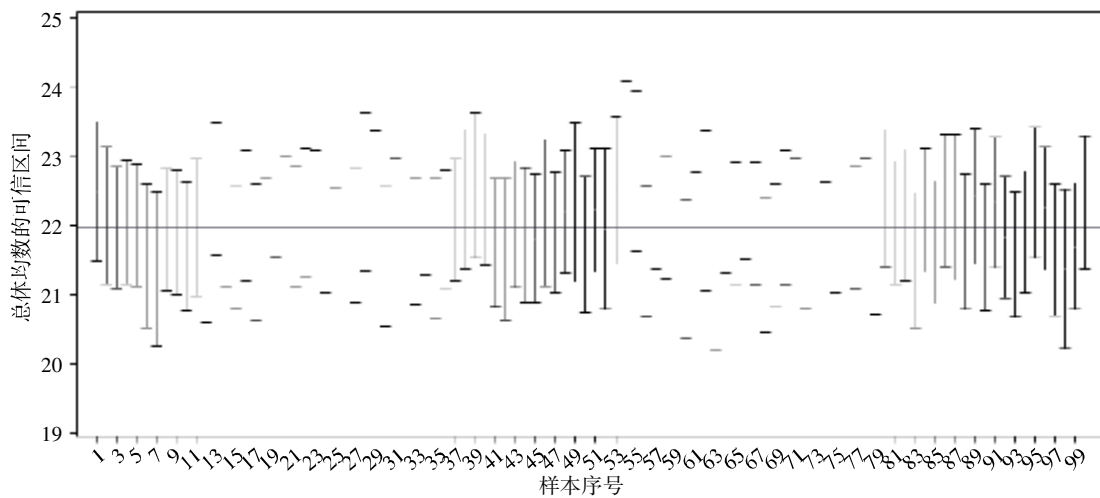


图 1 数量特征敏感问题加法模型分层三阶段抽样调查北京市 MSM 人群首次发生男男性行为年龄的计算机模拟结果  
Figure 1 Computer simulation result of the means of ages of the first occurrence of MSM by quantitative sensitive questions under stratified three-stage sampling survey

具有较高的效度;又因 100 个样本均数均接近同一数值(模拟总体均值),同时也说明本研究的分层三阶段抽样下数量特征敏感问题加法模型的调查方法及其统计公式具有良好的信度。

### 3 讨论

本团队对数量特征敏感问题整群抽样、二阶段抽样、分层两阶段整群抽样等复杂抽样的统计方法进行过研究<sup>[11-13]</sup>。在此研究基础上,本文对数量特征敏感问题加法模型下的分层三阶段抽样,从数学上首次推导出总体均数的估计量及其方差、估计方差的统计公式,具有统计理论方法学意义。

新的调查方法及其统计公式的效度与信度需采用大样本量下大个数样本的重复调查分析来验证,这对敏感问题的实例调查几乎不可能。本文采用计算机模拟的方法,不仅能实现理论上的大量重复抽样调查,还能保证获取资料的各项条件的稳定性<sup>[14]</sup>。蒙特卡洛方法(Monte Carlo method)是一种使用随机数(或伪随机数)来解决很多统计计算问题的统计模拟方法<sup>[15]</sup>,通过“实验”,以随机事件出现的频率估计其概率,或者得到这个随机变量的某些数字特征<sup>[16]</sup>。因 SAS 在进行大量重复抽样和蒙特卡洛研究的过程中效率更高<sup>[17]</sup>,本文基于 SAS 进行大样本个数(100 个样本)大样本量(平均每个样本包含 2 533 例模拟调查对象)的计算机模拟实验,对本研究的数量特征敏感问题加法模型下分层三阶段抽样的调查方法及其统计公式,得到了高信度、高效度的评价结论。运用本研究设计的调查方法及其公式,对北京市男男性

行为者这一性病、艾滋病高危人群的敏感问题调查分析,为有关部门制订性病、艾滋病预防控制规划、措施提供了科学依据,具有较大的实际意义。

荷兰 Utrecht 大学 Lensvelt-Mulders 等<sup>[18]</sup>收集了 1965 年 Warner 提出随机应答模型以来到 2000 年共 35 年间应用传统调查方法(包括面对面访谈、电话访谈、问卷调查等方法)与随机应答技术调查各类敏感性问题的调查结果,并进行了 Meta 分析。结果显示,对于敏感性特征已知的 6 项研究,随机应答技术与传统调查方法相比能显著提高正确应答率。对于敏感性特征未知的 32 项研究,采用随机应答技术所获得的结果其敏感性特征的比例或数量显著高于传统调查法,并且问题的敏感度越高,应用随机应答技术的效果就越好。该研究的结果充分证明应用随机应答技术调查敏感性问题的结果的准确性、可靠性方面较传统调查方法有着显著的优势。

本研究 100 个模拟样本得到的 100 个总体均数 95%可信区间全部包含模拟总体均数(真值),等价于进行了 100 个无统计学意义( $\alpha=0.05$ )的样本均数与总体均数比较的 Z 检验,说明 100 个 Z 值(样本均数与总体均数差之绝对值除以标准误的商)均小于  $Z_{0.05/2}=1.96$ ,较小的 Z 值并非由大的标准误造成,而是因为 100 个模拟样本均数的标准误均较小。由于篇幅限制,本文未包含分层三阶段抽样下数量特征敏感问题加法模型调查北京市 MSM 人群第 1 次发生男男性行为时年龄的计算机模拟结果,感兴趣的读者可以向作者联系。本文虽未对均数估计的无偏性进行数学证明,但 100 个模拟样

本的高效度评价结论也是对均数估计无偏性的较好实验证明。

[参考文献]

[1] 王云芳. 抽样调查技术在香港政府统计中的应用及启示[J]. 统计与管理, 2012(3): 88-89

[2] 陆守曾, 陈峰. 医学统计学[M]. 2 版. 北京: 中国统计出版社, 2007: 229-236

[3] 孙振球, 徐勇勇. 医学统计学[M]. 3 版. 北京: 人民卫生出版社, 2010: 500-506

[4] 孙萍, 赵东方. 敏感问题问卷调查 Warner 模型与卡片参数 P 的设计[J]. 统计与决策, 2012(9): 72-73

[5] 李伟. 数量特征敏感问题 RRT 模型整群抽样调查的统计方法及其效度信度模拟评价[D]. 苏州: 苏州大学, 2013

[6] 张倩倩, 还锡萍, 羊海涛, 等. 江苏省部分地区暗娼人数性病感染率及相关危险因素分析[J]. 南京医科大学学报: 自然科学版, 2012, 32(4): 473-478

[7] 于波, 高歌, 贺志龙, 等. 北京市 MSM 人群 RRT 模型的两阶段抽样调查[J]. 中国卫生统计, 2012, 29(3): 351-353

[8] 高歌, 金丕焕, 王旭辉. 分层三阶段抽样样本大小的研究及应用[J]. 中国卫生统计, 2000, 17(6): 325-327

[9] 王岩, 隋思涟, 王爱青. 数理统计与 MATLAB 工程数据分析[M]. 北京: 清华大学出版社, 2006: 9-11

[10] Wang J, Gao G. The estimation of sampling size in multi-stage sampling and its application in medical survey[J]. Applied Mathematics and Computation, 2006, 178(2):

239-249

[11] 王冕, 高歌. 数量特征敏感问题的整群抽样调查方法研究及应用[J]. 中国卫生统计, 2008, 25(6): 586-589

[12] 刘鹏, 高歌, 贺志龙, 等. 数量特征敏感问题加法模型二阶段抽样的统计方法及其应用[J]. 苏州大学学报: 医学版, 2011, 31(3): 384-387

[13] 朱宏儒, 高歌, 于明润, 等. 数量特征敏感问题分层两阶段整群抽样的统计方法[J]. 南京医科大学学报: 自然科学版, 2009, 29(6): 909-912

[14] Chen XY, Gao G, Yu MR. Stratified random sampling on simmons model for commercial sex survey among MSM in Beijing[A]//Zhu KL, Zhang H. Contemporary innovation and development in statistical science. Sydney: Aussino Academic Publishing House, 2012: 263-267

[15] Robert CP, Casella G. Monte Carlo statistical methods[M]. 2ed. New York: Springer, 2004: 77

[16] Du QQ, Gao G, Jin ZD. Application of monte carlo simulation in reliability and validity evaluation of two-stage cluster sampling on multinomial sensitive question[A]// in BX, Ma MD, Chang JC. Information computing and applications. Chengde: Springer, 2012: 261-268

[17] 谭志军, 徐勇勇, 曹文君, 等. 应用 SAS 和 SPSS 进行复杂抽样[J]. 中国卫生统计, 2012, 29(1): 145-146

[18] Lensvelt-Mulders, Gerty JLM, Hox L, et al. Meta-analysis of randomized response research: thirty-five years of validation [J]. Sociological Methods & Research, 2005, 33(3): 319-348

[收稿日期] 2014-01-08

## 科技出版物中阿拉伯数字的书写规则

1. 为使多位数字便于阅读, 可将数字分成组, 从小数点起, 向左或向右每 3 位分成 1 组, 组间留空隙(约为一个汉字的 1/4), 不得用逗号、圆点或其他方式。
2. 纯小数必须写出小数点前用以定位的“0”。
3. 阿拉伯数字不得与除万、亿及法定计量单位词头外的汉字数字连用。如 453 000 000 可写成 45 300 万或 4.53 亿或 4 亿 5 300 万, 但不能写成 4 亿 5 千 3 百万; 三千元写成 3 000 元或 0.3 万元, 但不能写成 3 千元。
4. 一个用阿拉伯数字书写的数值, 包括小数与百分数, 不能拆开转行。
5. 表示用阿拉伯数字书写的数值范围, 使用波浪号“~”。如 10%~20%,  $(2\sim6)\times 10^3$  或  $2\times 10^3\sim 6\times 10^3$ , 30~40 km。

(本刊编辑: 接雅俐)