文章编号: 1004-4353(2015)03-0238-06

基于分层三阶段抽样方法的大学生考试作弊次数调查与分析

周 霞
(阜阳师范学院 数学与统计学院, 安徽 阜阳 236037)

摘要: 针对大学生考试作弊的问题,对安徽省某地区某院校大学生的考试作弊次数进行了抽样调查分析. 为提高正确应答率和保护被调查者的隐私,使用随机应答技术(RRT)分层三阶段抽样方法和数量特征敏感问题乘法模型. 根据 Cochran 抽样理论,结合概率论与数理统计知识得出,该校学生在 30 科的考试中学生平均作弊次数为 12 次. 由此得知,治理考试舞弊、严肃考风考纪已成为高校教学的当务之急.

关键词: 分层三阶段抽样; 数量特征敏感乘法模型; 随机应答技术; 大学生考试作弊
中图分类号: O212.2 **文献标识码**: A

The survey and analysis of the number of undergraduates' exam cheating based on stratified three-stage sampling method

ZHOU Xia
(*School of Mathematics and Statistics*, *Fuyang Teachers College*, *Fuyang* 236037, *China*)

Abstract: Aim at the problem of the cheat in the examination of undergraduates, the number of exam cheating of undergraduates' in the Anhui province local normal colleges and universities is sampled and analyzed. To improve the correct response rate and protect the privacy of respondents, applying randomized response technique (RRT) stratified three-stage sampling method and quantitative characteristic sensitive multiplication model. According to the sampling theory of Cochran, combine with Probability and Mathematical Statistics, show that, the average number of exam cheating of undergraduates is 12 in 30 examination subjects. So, controlling cheating in the examination and solemnizing exam discipline have become urgent affairs.

Key words: stratified three-stage sampling; quantitative characteristic sensitive multiplication model; Randomized response technique (RRT); undergraduates' cheating in the examination

0 引言

考试是督促复习、巩固知识、评定学习效果的重要手段. 然而文献[1]表明,目前高校的考风状况不容乐观,作弊行为已经进入到一些学生的基本价值体系,形成一种"作弊文化". 2005 年,肖红新等[2]对某高校的考风考纪进行了随机抽样调查,结果显示: 20% 的学生表示经常参与作弊, 35.33% 的学生回答"经常看见考试作弊行为", 59.24% 的学生回答"偶尔看见", 5.43% 的学生回答"从没有看见". 文献[3]报道,某高校只有29.1%的学生回答"从未想过作弊". 文献[4]对甘肃农业大学学生进行了随机抽样调查,结果显示: 48.97% 的学生表示"偶尔作弊", 21.04% 的学生表示"经常作弊", 29.99% 的学生表示"从未作弊". 由此可知,

收稿日期: 2015-06-27
作者简介: 周霞(1981—),女,副教授,研究方向为随机时滞微分方程稳定性.

有作弊倾向或者作弊的高校大学生占相当大的比例，因此，高校整治考试舞弊行为势在必行.

二阶段或分层二阶段抽样方法[5-10]是目前大规模抽样调查敏感问题的常用方法. 近年来，三阶段抽样和分层三阶段抽样调查方法也偶有被运用于敏感问题调查[11-12]. 在相关模型研究中，无关联模型和加法模型研究得较多，而对数量特征敏感问题的乘法模型讨论得相对较少，其中：张振花[10]对二项选择敏感问题改进的 RRT 模型进行了二阶段抽样，并将结果运用于对大学生考试作弊行为所占的比例这一属性敏感的问题的调查中；范玉波[11]给出了数量特征敏感问题的无关联模型、加法模型和乘法模型的三阶段抽样和分层三阶段抽样方法下的样本量计算公式及估计量方差. 本文采用分层三阶段抽样方法，对安徽省某大学的大学生考试作弊次数这一数量特征敏感问题的乘法模型进行了研究，具体作法为：以性别作为分层依据，将该校 2015 届毕业生分为男生和女生两层；以该校的教学学院为一级单位，现有毕业生所分布的专业为二级单位，2015 届毕业生为三级单位，然后进行分层分阶段抽样调查，给出各层各阶段大学生考试作弊平均次数估计值.

1　大学生考试作弊次数的分层三阶段抽样调查

1.1　抽样总体、调查指标及调查目的

考虑到高校中各教学学院大一、大二、大三学生各学期考试科目门次不同及考试、考查课程分布的差异性，同时减少被调查者担心因暴露个人作弊行为而被处分的顾虑，该调查以安徽省某地区某高校的 2015 届毕业生作为抽样总体，其总体容量 $N=4\ 666$；以大四学生作弊经历的次数作为调查指标；通过调查，掌握高校学生考试作弊现状，为高校进一步分析作弊原因和制定相关对策提供参考.

1.2　分层三阶段抽样

按照性别将该校 2015 届毕业生分成两层（$H=2$），男生（$h=1$）和女生（$h=2$），其中男生 1 816 名，女生 2 850 名. 各层由 15 个教学学院（一级单位）组成（$N_h=15$，$h=1,2$），第 h 层第 i 个教学学院由 N_{hi} 个不同专业（二级单位）构成，第 h 层第 i 个教学学院第 j 个专业由 N_{hij} 个学生（三级单位）组成，具体情况见表 1 和表 2.

表 1　该校男生分布情况

学院	专业	男生人数	学院	专业	男生人数
文学院	汉语言文学	97	商学院	工商管理	63
	新闻学	16		财务管理	128
数学与统计学院	数学与应用数学	180		物流管理	37
	信息与计算科学	26		人力资源管理	18
	统计学	24	计算机与信息工程学院	计算机科学与技术（师范）	24
经济学院	经济学	48		计算机科学与技术	100
	电子商务	37		信息工程	61
物理与电子工程学院	物理学	63		信息管理与信息系统	25
	电子信息科学与技术	91	历史文化与旅游学院	地理科学	23
	科学教育	11		历史学	22
	应用物理学	34		资源环境与城乡规划管理	36
美术学院	绘画	7	体育学院	表演	8
	美术学	21		民族传统体育	27
	艺术设计	41		体育教育	167
化学与材料工程学院	材料化学	52	外国语学院	英语非师范	37
	化学	42		英语师范	6
	应用化学	44	音乐舞蹈学院	音乐表演	3
生命与食品工程学院	动物科学	10		音乐学	19
	生物科学	35	教育学院	小学教育	12
	园林	36		学期教育	1
政法学院	法学	69		应用心理学	9
	思想政治教育	6			

表2 该校女生分布情况

学院	专业	女生人数	学院	专业	女生人数
文学院	汉语言文学	250	商学院	工商管理	66
	新闻学	41		财务管理	238
数学与统计学院	数学与应用数学	266		物流管理	45
	信息与计算科学	20		人力资源管理	45
	统计学	24	计算机与信息工程学院	计算机科学与技术(师范)	25
经济学院	经济学	59		计算机科学与技术	71
	电子商务	52		信息工程	37
物理与电子工程学院	物理学	24		信息管理与信息系统	23
	电子信息科学与技术	54	历史文化与旅游学院	地理科学	42
	科学教育	27		历史学	35
	应用物理学	7		资源环境与城乡规划管理	37
美术学院	绘画	8	体育学院	表演	71
	美术学	48		民族传统体育	9
	艺术设计	58		体育教育	55
化学与材料工程学院	材料化学	45	外国语学院	英语非师范	313
	化学	66		英语师范	53
	应用化学	47	音乐舞蹈学院	音乐表演	14
生命与食品工程学院	动物科学	15		音乐学	71
	生物科学	68	教育学院	小学教育	104
	园林	30		学期教育	116
政法学院	法学	81		应用心理学	54
	思想政治教育	36			

第一阶段:从各层($h=1,2$)分别随机抽取7个教学学院作为一级单位,$n_h=7$,$h=1,2$.第二阶段:从h层的第i个被抽中的教学学院中再随机抽取n_{hi}个不同的专业作为二级单位.第三阶段:从h层第i个被抽中的教学学院中的第j个不同的专业里再随机抽取n_{hij}个三级单位.第一层被抽中的三级单位总数为660名,第二层被抽中的三级单位总数为850名.对每个被抽中的学生,采用数量特征敏感问题乘法模型进行调查,各层中被抽中的一级单位、二级单位、三级单位如表3和表4所示.

1.3 考试作弊次数的数量特征敏感问题乘法模型的随机应答技术

为了能更准确地调查出被抽中的1 510名学生在校期间的考试作弊次数,又能有效保护学生的隐私,本文应用数量特征敏感问题乘法模型,并采用随机应答技术进行调查.为此,设计了一套随机装置:在黑色的盒子中放置大小、质量、质感相同的10个小球,小球上分别贴有1、2、…、9、10的数字标签.每名被抽中的学生从黑色盒子中有放

表3 第一层中(男生)被抽中的一级单位、二级单位和三级单位的人数

学院	专业	人数	学院	专业	人数
文学院	汉语言文学	60	商学院	工商管理	50
数学与统计学院	数学与应用数学	120		财务管理	100
	统计学	10	计算机与信息工程学院	计算机科学与技术	75
经济学院	经济学	30		信息工程	50
物理与电子工程学院	物理学	50	历史文化与旅游学院	地理科学	15
	电子信息科学与技术	70		历史学	10
				资源环境与城乡规划管理	20

表 4　第二层中(女生)被抽中的一级单位、二级单位和三级单位的人数

学院	专业	人数	学院	专业	人数
文学院	汉语言文学	190	商学院	工商管理	45
				财务管理	180
数学与统计学院	数学与应用数学	220	计算机与信息工程学院	计算机科学与技术	50
	统计学	10		信息工程	15
经济学院	经济学	30	历史文化与旅游学院	地理科学	30
物理与电子工程学院	物理学	10		历史学	20
	电子信息科学与技术	30		资源环境与城乡规划管理	20

回地随机抽取一个小球，在确保无旁人知晓的情况下，将自己在校期间考试作弊次数和所抽到小球上的标签数字相乘，然后把乘后的结果写在调查表中.

1.4　学生考试作弊次数的调查结果

分别将第一层（男生）和第二层（女生）所填写的调查结果在 Excel 表格中进行汇总统计，并形成散点图，如图 1 和图 2 所示.

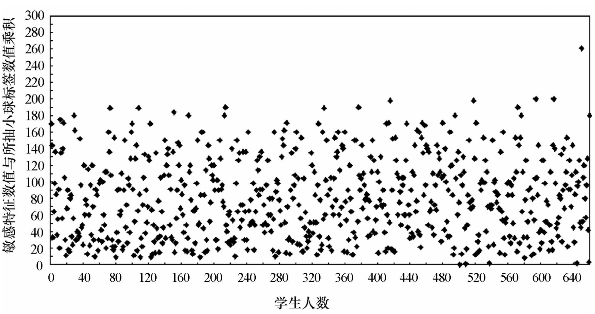

图 1　第一层(男生)调查结果的散点分布图

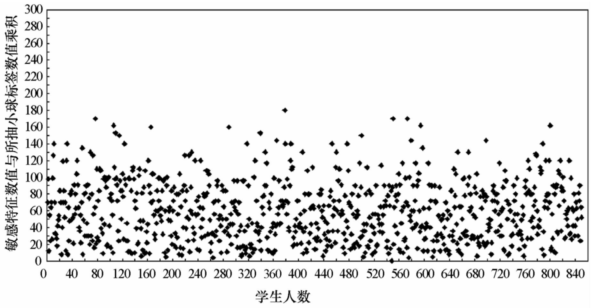

图 2　第二层(女生)调查结果的散点分布图

2　学生考试作弊次数的调查结果分析

2.1　各层各阶段考试作弊平均次数的估计量

设被抽中学生的考试作弊次数为随机变量 X，所抽小球上的标签数字为随机变量 Y，调查结果为随机变量 $Z(Z=XY)$. 设 $E(X)=\mu$，$\mathrm{Var}(X)=\sigma^2$，$\hat{\mu}$ 为 μ 的估计量，$\mathrm{Var}(\hat{\mu})$ 为总估计量 $\hat{\mu}$ 的方差. 记 x_{hijk} 为第 h 层第 i 个教学学院第 j 个专业中的第 k 个学生作弊次数，其中 $h=1,2$，$i=1,2,\cdots,7$，$j=1,2,\cdots,n_{hi}$，$k=1,2,\cdots,n_{hij}$；μ_{hij} 为 h 层第 i 个教学学院第 j 个不同专业学生考试作弊次数的平均值，$\hat{\mu}_{hij}$ 为 μ_{hij} 的估计量；μ_{hi} 为第 h 层第 i 个教学学院学生考试作弊次数的平均值，$\hat{\mu}_{hi}$ 为 μ_{hi} 的估计量；μ_h 为第 h 层学生作弊次数的平均值，$\hat{\mu}_h$ 为 μ_h 的估计量. 记 M_h 为 h 层含有的三级单位个数，这里 $M_1=1\,816$，$M_2=2\,850$. 根据概率论与数理统计知识可得：

$$\hat{\mu}_{hij}=\frac{1}{n_{hij}}\sum_{k=1}^{n_{hij}}x_{hijk}\,, \tag{1}$$

$$\hat{\mu}_{hi}=\frac{1}{\sum\limits_{j=1}^{n_{hi}}N_{hij}}\sum_{j=1}^{n_{hi}}N_{hij}\hat{\mu}_{hij}\,, \tag{2}$$

$$\hat{\mu}_h=\frac{N_h}{n_hM_h}\sum_{i=1}^{n_h}\frac{N_{hi}}{n_{hi}}\Big(\sum_{j=1}^{n_{hi}}N_{hij}\hat{\mu}_{hij}\Big), \tag{3}$$

$$\hat{\mu}=\frac{1}{N}\sum_{h=1}^{H}M_h\hat{\mu}_h=\sum_{h=1}^{H}\frac{M_h}{N}\left[\frac{N_h}{n_hM_h}\sum_{i=1}^{n_h}\frac{N_{hi}}{n_{hi}}\Big(\sum_{j=1}^{n_{hi}}N_{hij}\hat{\mu}_{hij}\Big)\right]. \tag{4}$$

2.2　被抽中的学生平均作弊次数的估计量

设 z_{hijk} 为第 h 层第 i 个一级单位内第 j 个二级单位里第 k 个三级单位（即调查对象）所填写调查表中的数值，其中 $k=1,2,\cdots,n_{hij}$. 由于使用随机装置得到的数据是每个被调查者敏感问题特征数值与其所抽中小球标签数字相乘的结果 z_{hijk}，并不是每个被调查对象的敏感问题特征数值 x_{hijk}，

故在式(1)中无法直接计算 $\hat{\mu}_{hij}$. 为求得 $\hat{\mu}_{hij}$,设抽中小球上的标签数字为随机变量 Y, Y 的取值为 $1,2,\cdots,10$,其均值为

$$\mu_Y=\frac{1}{10}(1+2+\cdots+10)=5.5. \tag{5}$$

这里分别记 $\mu_{z_{hij}}$, $\hat{\mu}_{z_{hij}}$, $s^2_{z_{hij}}$ 为 h 层第 i 个一级单位第 j 个二级单位所有回答数值的总体均值、样本均数和样本方差. 由于随机变量 X,Y 相互独立, $Z=XY$,于是可得

$$\mu_{z_{hij}}=\mu_{hij}\mu_Y, \tag{6}$$

$$\hat{\mu}_{hij}=\frac{\hat{\mu}_{z_{hij}}}{\mu_Y}=\frac{1}{\mu_Y}\hat{\mu}_{z_{hij}}=\frac{1}{\mu_Y}\sum_{k=1}^{n_{hij}}\frac{z_{hijk}}{n_{hij}}. \tag{7}$$

根据式(7)分别计算各层的 $\hat{\mu}_{hij}$,结果如表5所示. 由式(2)计算出各层的 $\hat{\mu}_{hi}$,结果如表6所示.

表5　各层中第 i 个一级单位内第 j 个二级单位敏感问题特征数值均值的估计量

h	i	j	$\hat{\mu}_{hij}$	h	i	j	$\hat{\mu}_{hij}$
1	1	1	13.384 8	2	1	1	11.825 8
	2	1	13.874 2		2	1	10.340 5
		2	13.963 6			2	9.000 0
	3	1	15.157 6		3	1	8.472 7
	4	1	7.236 4		4	1	13.290 9
		2	13.776 6			2	10.4
	5	1	15.309 1		5	1	9.078 8
		2	16.534 5			2	10.310 1
	6	1	12.501 9		6	1	9.909 1
		2	14.589 1			2	13.248 5
	7	1	15.200 0		7	1	12.381 8
		2	15.945 4			2	12.590 1
		3	15.818 2			3	10.027 3

表6　各层中第 i 个一级单位敏感问题特征数值均值的估计量

	$\hat{\mu}_{hi}$						
	$i=1$	$i=2$	$i=3$	$i=4$	$i=5$	$i=6$	$i=7$
$h=1$	13.384 8	13.884 7	15.157 6	11.100 3	16.130 3	13.297 2	15.677 2
$h=2$	11.825 8	10.229 6	8.472 7	11.289 5	10.042 8	11.053 2	11.656 7

由式(3)可得:

$$\hat{\mu}_1=\frac{1}{1\,164}[13.384\,8\times(97+16)+13.884\,7\times(180+26+24)+15.157\,6\times(48+37)+11.100\,3\times(63+91+11+34)+16.1303\times(63+128+37+18)+13.2972\times(24+100+61+25)+15.677\,2\times(23+22+36)]=13.946\,4;$$

$$\hat{\mu}_2=\frac{1}{1\,488}[11.825\,8\times(250+41)+10.229\,6\times(266+20+24)+8.472\,7\times(59+52)+11.289\,5\times(24+54+27+7)+10.0428\times(66+238+45+45)+11.053\,2\times(25+71+37+23)+11.656\,7\times(42+35+37)]=10.636\,7.$$

由式(4)可得

$$\hat{\mu}=\frac{1}{N}\sum_{h=1}^{H}M_h\hat{\mu}_h=\frac{1}{4\,666}(1\,816\times\hat{\mu}_1+2\,850\times\hat{\mu}_2)=11.924\,8.$$

3 结论与分析

本文采用随机应答技术，改进文献[11-12]中乘法模型的随机装置，对大学生“考试作弊次数”进行分层三阶段抽样调查，既很好地保护了被调查者的隐私，又提高了回答准确率. 调查结果显示：学生在校期间，30 科考试中学生平均作弊次数为 12 次；男生平均作弊次数是 14 次，女生平均作弊次数是 11 次；所抽中的 7 个教学学院中男生的平均作弊次数分别为 13、14、15、11、16、13、16 次，所抽中的 7 个教学学院中女生的平均作弊次数分别为 12、10、8、11、10、11、12 次；第一层（男生）中所抽中的来自抽中教学学院中被抽中专业的平均作弊次数分别为 13、14、14、15、8、14、15、17、13、15、15、16、16 次，第二层（女生）中所抽中的来自抽中教学学院中被抽中专业的平均作弊次数分别为 12、10、9、8、13、10、9、10、10、13、12、13、10 次. 由此可见，男生几乎每两门考试课程中就有一门出现舞弊行为，女生几乎 3 门考试课程中就有一门出现舞弊行为. 这说明，大学生考试作弊这一问题已经到了很严重的程度，必须加以充分重视并制定相应的对策.

参考文献：

[1] 陈富. 大学生考试作弊行为的实证研究[J]. 考试研究，2014，43(2)：73-82.

[2] 肖红新，梅景良，陈鑫珠，等. 大学生考试作弊情况调查分析及其遏制对策探讨[J]. 福建农林大学学报（哲学社会科学版），2009，12(1)：102-105.

[3] 李海. 大学生考试作弊情况调查分析及其遏制对策[J]. 调查与研究，2011，7(6)：169-170.

[4] 路建龙，逄蕾，王银. 大学生考试作弊现状调查及对策浅析[J]. 价值工程，2015，7：237-238.

[5] 濮翔科. 敏感问题 9 种 RRT 模型下（分层）二阶段抽样调查设计的统计方法及其应用[D]. 苏州：苏州大学医学部公共卫生学院，2012.

[6] Cruyff M J, Van Den Hout A, Van Der Heijden P G, et al. Log-linear randomized-response models taking self-protective response behavior into account[J]. Sociological Methods & Research, 2007, 36(2)：266-282.

[7] 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)：384-408.

[8] 周云华，高歌，濮翔科，等. 数量特征敏感问题加法模型分层二阶段抽样样本大小研究及其应用[J]. 中国卫生统计，2014，31(1)：45-48.

[9] Wang Jianfeng, Gao Ge, Fan Yubo, et al. The estimation of sample size in multi-stag esampling and its application in medical survey[J]. Applied Mathematics and Computation, 2006, 178：239-249.

[10] 张振花. 改进 RRT 模型的分层整群抽样在大学生考试作弊行为中的应用[D]. 兰州：兰州大学数学与统计学院，2013.

[11] 范玉波. 敏感问题 RRT 模型下（分层）三阶段抽样调查的统计方法及其应用[D]. 苏州：苏州大学医学部公共卫生学院，2013.

[12] 陈科锦，高歌，范玉波，等. 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[J]. 南京医科大学学报（自然科学版），2014，34(8)：1115-1119.