

中国人民大学

Renmin University
of China



性社会学研究所

Institute for Research on
Sexuality and Gender

<http://www.sexstudy.org>

[首页](#) | [学科建设](#) | [基础研究](#) | [应用研究](#) | [著作发布](#) | [图书检索](#) | [通俗文章](#) | [文献介绍](#) | [讨论地带](#) | [通讯文档](#)

您的位置: [首页](#) -> 专著《中国当代大学生的性观念与性行为》

《中国当代大学生的性观念与性行为》03

作者: 潘绥铭、曾静 来源: 商务印书馆, 2000年3月出版 类别: 专著《中国当代大学生的性观念与性行为》 日期: 2005.02.23 今日/总浏览: 1/2075

第一部分·第二章

我们调查了什么人

——随机抽样的方法与过程

一、调查总体的确定

我们的调查总体是全国所有向社会公开的正规高等院校的所有本科生。这个定义, 从3个方面对我们的调查总体作出了限定:

1、在学校类型方面, 我们调查的是那些由国家教委、国家各部委和各级地方政府的教育管理部门自己开办并且直接管理的高等院校, 也就是通常所说的公立正规高校。这就在所有者和办学途径的层次上, 排除了那些私立的、民办的、单位或者社团开办的、社会集资开办的各类高等院校。

同时, 我国还有一些军事方面的院校也属于国家正规高校, 但是由于它们的情况是不允许向社会公开的, 所以也被排除在我们的调查总体之外。

2、在学生来源方面, 我们调查的是那些通过国家正式高考(不包括成人高考)而进入高校的、脱产的、全日制的本科生, 也就是人们开玩笑说“有正式户口”的那些正规本科生。这样, 在学生资格的层次上, 我们排除了所有那些“五大”(广播电视大学、函授大学、夜大学、业余大学、职工大学)、成人教育、干部(教师)进修、未经高考而争取获得本科同等学历的各级各类高等院校。

3、我们调查的本科生是指: 学制为4年或者5年的、最终可以获得国家承认的本科学历的(“有全国粮票”的)大学生。在医学院校或者外加1年军训的院校里, 也包括5年级以上的、但仍然只能获得本科学历的那些大学生。这样, 即使在一个符合上述条件的正规高校的正规学生里, 也排除了所有的大专生、“先专后本”生、进修生、旁听生等等。

这样来确定我们的调查总体, 并不意味着对任何其它类型的高校和高校学生的任何贬低或者歧视。我们主要是从这样3个方面来考虑, 才做出如此限定的:

1、按照国家的有关规定, 凡是通过高考(不包括成人高考)而进入正规高等院校的本科生, 入学前和在校期间, 是

绝对不允许结婚的。但是其它各类高校和其它类的大学生，往往并没有这样的硬性规定。这就从性质上把两类大学生根本地区分开了。我们要调查的是性观念与性行为，但是可以结婚的人与不允许结婚的人，在这方面并不具备相同的基础与客观条件，因此不能放在同一个调查总体之内。

2、在中国目前的社会里，通常所说的“正规本科生”，是一个很特殊的群体。由于我们的正规高考制度非常严格，形成了“独木桥”和“跳龙门”的独特局面，因此所有那些通过正规高考而进入正规高校的人们，一直被全社会视为“骄子”、“宠儿”、“精华”。他们与其它各类大学生相比，不仅将来的前途很可能要好得多，而且在目前的地位、声望以及对于社会的影响力等等，都存在着极大的区别，有些情况下甚至是本质的区别。

在许多发达国家里，一个人上的是什么样的大学，往往仅仅表明此人受的大学教育是不是足够好；但是在目前中国，是不是正规本科生，却常常具有着受教育水平以外的许许多多其它意义，甚至是高人一等还是低人一等的大问题。所以我们应该把这两类人放在同一个总体之内。

3、从在校期间的生活来看，正规高校的大学生几乎都是全日制和寄宿制，走读的微乎其微，可以忽略不计。但是其它类型的高校大多数并不是这样的，至少并不象正规高校那样一律和严格。这个区别很重要。正是因为存在着一律的和严格的全日制和寄宿制，正规高校里才可能建立起一种独特的校园文化和校园生活方式，才可能使正规本科生在这样的熏陶中，产生出那些在全国所有正规高校里如出一辙的特质和特征，即人们通常所说的“书生气”。

且不论我们如何评价这种书生气，它不仅可以把正规本科生与社会一般青年相当严格地区别开来，而且还足以把正规本科生与其它各类大学生区别开来。

从社会化的角度来看，这种书生气所具有的巨大意义，对于一个人甚至对于整个本科生群体来说，往往并不亚于他们所获得的知识。其它类型高校里的学生，例如各种业余大学，尽管也可以获得本科学历甚至正规的本科毕业证书，但是由于他们没有校园生活的经历，会在许多重要方面与正规本科生出现巨大差异。因此，这两类大学生不应该放在同一个调查总体之内。

二、设计抽样方案

由于本课题的调查对象，如前面所述是一个庞大且分散的总体，目前还找不到这样一个包含了符合条件的所有在校本科生名单的现存资料，因此无法直接抽取所需的样本；而且我们的调查对象，随着升学、毕业、流失等情况的发生，每年在不断变化，若要收集名单，既费时耗力，还不准确，在课题经费紧张，时间有限的情况下，根本不可能实现。所以我们决定运用多级抽样，即通过抽取若干级中间单位，来达到抽取基本调查单位的目的。

在本次调查中，我们计划进行三级抽样。以全国所有的四年制及以上的正规高等院校为一级抽样框，在其中抽取若干学校；再以抽中的学校为二级抽样总体，抽取一定数量的系，对抽中的系发信，索要该系所有在校本科生的名单；最后以提供的学生名单为第三级抽样框，抽取部分学生，发放邮寄问卷进行调查。

当然，我们的抽样方案要遵循随机抽样的原则，保证总体中所有个体被抽中的机会都是均等的，没有根据任何主管意图来挑选或确定调查单位；同时要实现抽样效果最佳原则。即在固定的费用下，选取抽样误差最小的方案，或在要求的精度下，做到调查费用最少；最后还要考虑本次调查内容的敏感性以及邮寄调查的特点所造成的邮寄损失、应答率低、回收率低等情况。所以虽然采用多级抽样，但每一级抽样中又根据实际需要结合其它的抽样形式。

三、样本含量的确定

由于我们将抽样总体按学校的隶属关系和所在地的经济发展程度分成了20个单元，每个单元在总体中所占的平均比例为5%，所以我们将样本比例与总体比例的绝对差值定在 $5\% \div 2 = 2.5\%$ 。因此，在置信度为 $1 - \alpha = 95\%$ ，最在允许误差 $\Delta p = 2.5\%$ 的条件下，其最小保守样本含量（即有限总体的无放回抽样的最小样本含量） n_0 为：

$$n_0 = (t_{\alpha} / 2 / 2\Delta p)^2 = 1537$$

根据我们前两次所作的同类调查的55%和56%的问卷回收率，将这次调查的回收率保守估计为50%，所以应发出3074份邮寄问卷，取其整数，计划发出3000份问卷，由于在抄写过程中的疏漏和邮政分检中的机械操损伤，实际发出了2936份问卷，收回1921份，超出了计划最小样本含量。

四、抽样的具体实施过程

1、第一级抽样

根据1997年国家教委最新出版的《中国高等院校信息库》光盘资料信息，再参照《高校招生录取及填报志愿指南》（航空工业出版社，1997/3/1）收集的截至1996年底的高校信息，我们获得符合条件的全国高校的总体情况，共638所高校，分布在全国31个省、自治区、直辖市。本科生总数为1425600人。以这638所正规高等院校为第一级抽样框，从中抽取一定数量的学校。

考虑到作为抽样单元的单个学校的差异较大：首先，各个学校的大小规模很不一致，有的学校本科生人数达一万以上，而有的大学本科生人数还不足一千，因此各个学校在总体中所占的比重很不相同；其次，不同学校由于隶属关系的不同而有等级之分，致使调查对象在不同学校间差异大，一般重点大学录取分数高，一般院校录取分数低，所以学生在高考录取时就以高考成绩的高低以及各校录取分数的差异而有分流，而且一般省属院校以及地方所属院校对本地招生多，学生多居住在该省市附近；此外，由于各校的历史长短等因素，而有不同的校园文化，这会影响到学生的观念及行为；当然，不同学校位于不同地区，学生也会受当地的社区环境、文化经济的影响等种种因素，我们结合了分层等距抽样（stratified sampling）方法。

即先按某些特征将总体划分成有限个互不重叠的部分，每一个部分叫做一个子总体，再从每个子总体中抽取样本。它的好处是可以将一个内部差异很大的总体分成一些内部比较相似的子总体。这样由于子总体内的抽样单元间的差异可能较小，因此分层后的子样本可能具有较好的均匀性，有可能得到较高精度的估计量；其次在总体中常有一些比较特殊的单元，由于他们所占的比例很小，在简单随机抽样中往往会因为这些特殊单元的比例过高或过低而影响到调查结果，而分层抽样可将这些特殊单元当作一层对待，从而样本就更有代表性，最后，分层抽样从管理实施来说也是更为方便的。

由于应按各层间的差距尽可能大的原则来分层。而且各层间不会相互重叠。一般用于分层的理想标志是在调查中要加以测量的变量，所以依照前面所述的原因，我们选取两个变量对总体进行分层，第一个变量为学校的隶属关系，将总体分为4类：教委直属、部委直属、省属和地方政府管辖。第二个变量为学校的所在地，按照经济发达程度，将这些学校所在的省区从1—32编号，且将其分为5组（我们将深圳算作独立的省区也列在内），其中A组：1—7 包括1上海、2北京、3天津、4重庆、5深圳、6广东、7海南；B组：8—10 包括8福建、9浙江、10江苏；C组：11—17 包括11四川、12湖北、13山东、14辽宁、15黑龙江、16吉林、17河北；D组：18—23 包括18河南、19湖南、20安徽、21江西、22山西、23广西；E组：24—32 包括24陕西、25云南、26内蒙古、27贵州、28宁夏、29甘肃、30新疆、31青海、32西藏。由此，将总体分为4×5共20个子总体。

具体操作如下：将高校情况录入计算机，在excel表单里，先将高校按教委直属、部委直属、省属、地方所属先后排序，再将每一类里的高校按其所在省区依照ABCDE的先后排序，最后我们得到一个包含一级抽样单位的抽样框，由于教委直属D组这一个子类的学校数为0，实际得到19层。在计算机里的排列如下表所示：（因为表格太长，无法全部列出，仅以教委属学校为例）

学校名	所在省区	类别号	编号	隶属关系	类型编号	本科生数
上海交通大学	上海	A	1	教委直属	1	9702
复旦大学	上海	A	1	教委直属	1	5445
同济大学	上海	A	1	教委直属	1	7137
华东师范大学	上海	A	1	教委直属	1	5379
上海外国语大学	上海	A	1	教委直属	1	1910
华东理工大学	上海	A	1	教委直属	1	6280
北京大学	北京	A	2	教委直属	1	6774
北京师范大学	北京	A	2	教委直属	1	4625
清华大学	北京	A	2	教委直属	1	10061
北京外国语大学	北京	A	2	教委直属	1	1400
国际关系学院	北京	A	2	教委直属	1	799
中国人民大学	北京	A	2	教委直属	1	5414
北京语言学院	北京	A	2	教委直属	1	478
天津大学	天津	A	3	教委直属	1	8259
南开大学	天津	A	3	教委直属	1	6213
西南师范大学	重庆	A	4	教委直属	1	4474
重庆大学	重庆	A	4	教委直属	1	7033
中山大学	广东	A	6	教委直属	1	5899
华南理工大学	广东	A	6	教委直属	1	8419
广州外国语学院	广东	A	6	教委直属	1	1320
厦门大学	福建	B	8	教委直属	1	6939
浙江大学	浙江	B	9	教委直属	1	8844

南京大学	江苏	B	10	教委直属	1	6630
东南大学	江苏	B	10	教委直属	1	6834
四川联合大学	四川	C	11	教委直属	1	11913
华中师范大学	湖北	C	12	教委直属	1	4779
武汉大学	湖北	C	12	教委直属	1	7677
华中理工大学	湖北	C	12	教委直属	1	9461
山东大学	山东	C	13	教委直属	1	6055
青岛海洋大学	山东	C	13	教委直属	1	2838
大连理工大学	辽宁	C	14	教委直属	1	7168
东北师范大学	吉林	C	16	教委直属	1	4723
吉林大学	吉林	C	16	教委直属	1	5191
西安交通大学	陕西	E	24	教委直属	1	8292
西北大学	陕西	E	24	教委直属	3	4931
陕西师范大学	陕西	E	24	教委直属	1	4342
兰州大学	甘肃	E	29	教委直属	1	5702

在第一阶段抽样上，应通过提高抽样比来提高精度，所以在638所符合条件的高校里，计划抽取150所学校。

实际操作如下：先产生一个随机数，作为在抽样框里的抽样起点；抽样间距即 $[638 \div 150] = 5$ 。然后，从起点开始，以5的抽样间距等距抽取学校。

抽样结果见下表：

隶属关系	类	校数	本科生总数	抽校数	抽校本科生数	抽中系数
教委属	A	20	107021	6	38164	37
	B	4	29247	1	8844	8
	C	9	59805	3	24306	24
	E	4	23267	1	4931	5
部属	A	80	146718	20	35801	48
	B	28	63034	9	25377	27
	C	85	216367	17	47001	47
	D	42	95176	8	21521	22
	E	32	70847	8	21697	23
省属	A	55	107720	12	28822	35
	B	35	69354	9	20196	24
	C	100	181717	23	49842	57
	D	58	133471	15	43624	47
	E	67	101151	15	31087	38
地方属	A	6	8010	1	1453	2
	B	4	2968	1	781	1
	C	7	9055	1	117	1
	D	1	340			
	E	1	332			
总计		638	1425600	150	403564	446

2、第二级抽样

这一级抽样采用配额抽样方法。我们计划平均每所学校抽3个系（学院）¹，共450个系。由于各校的规模不一，有的学校多达29个系，有的学校只有1个系，为了增加样本的覆盖率，我们以学校所有的系（学院）数进行配额。但由于我们掌握的资料中有很大部分的学校没有提供系数，所以只得通过学校的本科生数推断系数，假设本科生人数多的学校其系数也多。我们随机抽取30个注明了系数的学校，计算它们的本科生数和系（学院）数的皮尔逊相关系数，结果为0.826，这证明我们的推断是合理的。

所以我们按照如下方法抽取系：

$$d = [(x_i \div x_{\text{平}}) \times 3]$$

d：抽取系数

x_i ：抽中学校本科生数， $i = (1, 2, 3, \dots, 150)$

$x_{\text{平}}$ ：各校平均本科生数

在具体抽取各系时，由于不能获取各系名单，我们在第一级抽样框中得到的实际为该学校的各专业名，对其进行等距抽样，以学校的邮政编码的第一位数字作为抽样起点，同时经计算得出来平均每校有11个专业，因此抽样距离为 $[11 \div 3] = 4$ ，通过抽得的专业名在国家教委的专业设置目录上查找其所属的系，最后实际共抽得446个系，给这些系的教务秘书发出信，恳请提供该系所有在校男女本科生的名单。

第一级抽样结果的汇总如下表所示：

隶属关系	类	校数	抽校数	抽中系数
教	A	20	6	37
	B	4	1	8
	C	9	3	24
	E	4	1	5
	部	A	80	20
B		28	9	27
C		85	17	47
D		42	8	22
E		32	8	23
省	A	55	12	35
	B	35	9	24
	C	100	23	57
	D	58	15	47
	E	67	15	38
地	A	6	1	2
	B	4	1	1
	C	7	1	1
	D	1		
	E	1		
总计		638	150	446

3、第三级抽样

在这一级抽样中采取等距抽样方法。由于我们的调查时间安排得紧凑，且考虑到各校回信时间不一，若要等齐所有回信，又恐学生外出实习，毕业或放暑假，不能收到调查问卷。因此，截至4月4日，我们共收到116封回信，寄来的学生名单总人数超过30,000，于是决定每15人抽取1名学生，按此方法实际抽得2001人（实际寄1976封平信，4月9日发

出)，后又收到寄回的名单，还是按照相同的间距抽取141人（寄141封挂号信，4月16日发出，同时寄出上次信件的催信）。至此，我们共收信122封，名单总人数31,875，抽得2181人。这次抽样对名单没有做任何形式的排序，因为我们没有收齐所有的回信，无法估计将会有名单总数，当然，有的学校由于种种原因根本不愿寄来学生名单，所以，我们假设在这些回来的学生名单数中等距抽中的19类人数在总样本量中的比例，与19类的本科生总数占总体的比例是一致的。

后来又陆续收到一些寄回的学生名单，到4月下旬估计不会再有名单寄回时，我们将19类抽中数比例与19类总体数比例做了比较，发现有几类出入较大，为此，我们对教委直属A类和部委直属A类进行追要本科生名单。最终，返回信的学校为81所，系数133个，回信系所提供的本科生名单总数为37,644。因为我们原定的样本量为3000，所以还需要做调整补抽，使比例一致，才能反映总体情况。我们按分层等距随机抽样出的150所学校，形成了一个能代表638所总体的子总体，所以每一类学校应抽取人数为：

$$\text{抽取的人数} = 3000 \times (\text{抽中的各类学校本科生数} \div \text{抽中学校的总共本科生数})$$

$$\text{补抽人数} = \text{应抽人数} - \text{实抽人数}$$

$$\text{补抽间距} = [(\text{每类学校名单人数} - \text{已抽人数}) \div \text{补抽人数}]$$

共补抽619人（4月24日寄出828封挂号信，141封催信，4月30日寄出催信）。由于在这之前有的类抽中人数已超过了应抽人数，所以我们实际抽中2936人。各类实际抽中人数的比例与理想比例是相当接近的。

隶属	类	回信校	回信系	名单人数	男	女	实抽人数	实抽比例	抽男	抽女
教	A	4	9	4633	2181	2451	257	8.753	127	130
	B	1	1	636	527	109	59	2.009	49	10
	C	3	8	1951	1311	630	176	5.994	119	57
	E	1	1	328	192	116	34	1.158	23	11
部	A	6	11	3301	2147	1154	247	8.412	167	80
	B	4	6	1564	1331	233	174	5.926	152	22
	C	6	9	3080	2399	700	336	11.44	267	69
	D	5	5	1666	1260	396	148	5.040	116	32
	E	4	8	3001	2286	715	201	6.846	155	46
省	A	6	12	2642	1420	1222	199	6.777	106	93
	B	5	8	1560	1105	455	140	4.768	93	47
	C	16	23	6437	4082	2375	428	14.57	273	155
	D	11	20	4641	3213	1428	307	10.45	195	112
	E	7	10	1988	1245	743	215	7.322	129	86
地	A	0	0	0	0	0	0	0	0	0
	B	1	1	176	74	102	12	0.408	5	7
	C	1	1	40	24	16	3	0.102	2	1
	D									
	E									
总计		81	133	37644	24797	12845	2936		1978	958

4、说明

需要说明的是本次调查对象为大学生群体，同质性高，因此抽样时未考虑其它因素。同时，因为不断的院系调整，原来抽中的学校由于合并或隶属关系转移有的升级，有的降级，尽管我们参考了各方资料，仍然会有遗漏和不准确，造成抽样的偏差；由于考虑到军队院校的管理特点，我们还人为排除了军队院校；再有，因为我们是先对抽中的系发信，索要该系所有在校本科生的名单，然后根据名单上提供的信息，抽取学生，在这一过程中，被抽中的系并未全部寄来回

信，尽管我们的抽样汇总表显示出各组类占总体比例一致，但在每一组类里，可能造成某些回信集中在某些省区，使得被抽中的学生在全国省区中分布不均；而且有的系即便返回信，但提供的本科生名单也并不完全，有的人为漏掉了毕业生。这些非抽样误差，是很难避免的，但在以后的调查中，我们必须寻求有效的措施尽量减少它们的出现，以求提高抽样精度，为调查打下良好的基础。

1 我们在这里所指的学院是系级学院，其下不隶属有系级单位。

关闭窗口

文章版权归原作者所有，未经允许请勿转载，如有任何问题请联系我们。