

# 出生性别比因素的均匀抽样探索

钱 进

杨小莉

(中南财经政法大学信息学院) (武汉市武昌区妇幼保健所)

**[摘要]** 为了把均匀设计用于抽取数据库中数据的一个子集,特寻求一个实际用例,探索可能性或筛选效率。正好武汉市武昌区,为了弄清出生人口性别比持续偏高的缘由,动用了该区妇幼保健所积累一年多的“新生儿出生医学记录”数据库进行多属性分析。于是,根据全部数据结果的启示,选择 37 行 10 列的均匀设计表,依据“直方图变尺度”安排因素和水平,通过数据库结构查询功能,选出 37 个记录。日后经适当的分析方法,检验样本结果和全体数据分析的结果有无显著性差异。本文旨在摸索实际的操作过程,发现问题,启发应用;还远远不能进入理论推导阶段。

**[关键词]** 均匀设计, 出生人口性别比, 数据库, 变尺度

## 一、 数据库及其背景

近一年来,和全国大多数地区相似,武汉市武昌区出生人口性别比持续偏高,引起社会各界的高度重视。为了弄清原由,找出问题的症结,该区妇幼保健所对所办理的 3682 份“新生儿出生医学记录”进行多属性分析。结果显示,该区最近一年的出生人口性别比为 122.75,超过了国际公认的 102—107 的标准,经卡方检验出生人口性别比与产次和出生体重有显著性差异。提出建议:1、应多方面寻找出生人口性别比过大的原因,跳出思维定式;2、传统的性别偏好思想仍很严重,需各个层面加大宣传力度,转变育龄群众的生育观念:保持合理的人口出生性别结构、建设新型生育文化,是维护国家、民族利益的大事<sup>[1]</sup>,采取措施,指出滥用 B 超检查、人工终止妊娠以满足个人利益而选择胎儿性别是违法行为,将被严肃查处。3、各级政府要采用系统工程方法,综合治理。法规,政策的配套完善是解决出生人口性别比失调的重要手段。<sup>[1]</sup>

### 1、资料与方法

1.1 资料来源:2002 年 7 月 1 日至 2003 年 6 月 30 日出生、且户口上在武昌区新生儿的“新生儿出生医学记录”。“新生儿出生医学记录”均由武汉市各大医院接生医生负责填写。

1.2 指标选择:从“新生儿出生医学记录”中选择了 10 个重要属性作指标,从已录入的 EXCEL 报表数据直接导入 VISUAL FOXPRO 6.0 数据库表中,并投影到几个对性别有影响的属性(参见表项)上。

### 2、性别比概念与标准

2.1 所谓某人口出生性别比,是指该人口某一时期(通常为一年)内出生的男婴总数与女婴总数的比值,用每百名出生女婴数相对应的出生男婴数表示。<sup>[2]</sup>

2.2 值域标准:20 世纪 50 年代中期,联合国认为:“男婴出生数置于 102~

107 之间。<sup>[2]</sup> ”一直被国际社会公认为通常理论值，其他值域则被视为异常。

### 3、九因素对性别比影响的数据分析

如果把新生儿属性中的性别看作主要考察对象，上述其余 9 个属性视为影响因素，通过数据库的投影运算和连接，用统计命令计算出属性值下的出生性别比。所得计算结果如表 1 至表 10 所示：

表 1 武昌区近一年出生人口性别比

总计	男孩数	女孩数	性别比
3682	2029	1653	122.75

表 3 孕周与出生性别的关系

行次	孕周	男孩数	女孩数	性别比
1	37 周以下	78	45	173.33
2	37 周至 41 周	1910	1583	120.73
3	42 周及以上	41	26	157.69

表 4 出生体重与出生性别的关系

行次	体重	男孩数	女孩数	性别比
1	2500g 及以下	50	44	113.64
2	2500g-4000g	1774	1513	117.25
3	4000g 以上	205	96	213.54

$X^2=22.4163$  P 0.001 有极显著性差异

表 2 出生月份与出生性别的关系

行次	月份	男孩数	女孩数	性别比
1	02 . 7	124	112	110.71
2	8	227	199	114.07
3	9	213	144	147.92
4	10	222	196	113.27
5	11	219	179	122.35
6	12	216	178	121.35
7	03 . 1	206	161	127.95
8	2	161	129	124.81
9	3	184	141	130.5
10	4	95	66	143.94
11	5	29	29	100
12	6	133	119	111.76

表 5 母亲年龄与出生性别的关系

行次	母亲年龄	男孩数	女孩数	性别比
1	20 岁及以下	17	6	283.33
2	21-25 岁	584	493	118.46
3	26-30 岁	1177	936	125.75
4	31-35 岁	226	194	116.49
5	36 岁及以上	25	24	104.17

表 6 母亲文化与出生性别的关系

行次	母亲文化	男孩数	女孩数	性别比
1	无记录	1026	828	123.91
2	大学及以上	461	368	125.27
3	中学文化程度	537	454	118.28
4	小学及以下	5	3	166.67

表 7 父亲年龄与出生性别的关系

行次	父亲年龄	男孩数	女孩数	性别比
1	25岁及以下	146	104	140.38
2	26-30岁	1008	786	128.24
3	31-35岁	696	601	115.81
4	36岁及以上	179	162	110.49

表 8 父亲文化与出生性别的关系

行次	父亲文化	男孩数	女孩数	性别比
1	无记录	1027	828	124.03
2	大学及以上	520	412	126.21
3	中学文化程度	478	411	116.3
4	小学及以下	3	2	150

表 9 孕次与出生性别的关系

行次	孕次	男孩数	女孩数	性别比
1	1	1331	1048	127
2	2	440	366	120.22
3	3	164	164	100
4	4	66	48	137.5
5	5	19	16	118.75
6	6	7	8	87.5
7	7	1	2	50
8	8	1	1	100

表 10 产次与出生性别的关系

行次	产次	男孩数	女孩数	性别比
1	1	1939	1607	120.66
2	2	87	43	202.33
3	3	3	3	100

$X^2=7.6597$  P 0.05 有显著性差异

#### 4、讨论（摘要）

从表 9、表 10 中可以看出，第一次生产的新生儿出生数占新生儿总数的 96.31%，其性别比为 120.66，其中孕 1 产 1 的新生儿出生数占第一次生产的 67.03%，占新生儿总数 64.61%，性别比为 127.00。则提示我们：武昌区出生人口性别比的偏高，不能简单地归咎于进行了“非医学鉴定胎儿性别和选择性别终止妊娠”。

这也例证了马瀛通等在《中国出生人口性别比研究》论文中所归纳的六个“重新认识”的问题<sup>[3] [4]</sup>，的确有必要重新认识。影响出生性别比高于 107 的因素中，有些至今仍未被人们所认识。在该区近 100 万人口中，大专院校、省级机关、文化卫生部门的高知人群占极大比重，出生性别比 120 是否一定不正常？表 11 37 行 10 因素均匀表  $U_{37}(12^{10})$

而占总数 3.53%的产 2 新生儿,性别比有近 80 个百分点的递增,证明了较高孩次出生性别比受男性偏好影响较严重,要引起重视。出生性别比的异常升高,若持续下去,将对人口结构产生重大负面影响,对未来社会的稳定运行,社会伦理道德体系也会造成一定冲击。同时男女失调也将带来一系列的社会问题,如婚姻性别挤压问题,对家庭和社区的冲击问题,就业性别挤压问题等。出生性别比升高及其可能带来的社会问题,是需要认真解决的。

## 二、均匀设计

### 1、选择均匀表

根据样本为总体的百分之一取表行。生成向量用<sup>[5]</sup>中第 12 页随机排列表第 19 行前 10 个数: 9, 4, 3, 1, 2, 8, 6, 7, 0, 5; 将零转换为从 1 起始。生成的 37 行 10 因素均匀设计表如表 11 中  $U_{37}(12^{10})$  所示。

### 2、生成因素水平安排表

第 1 列以 1-20 安排男孩, 21-27 安排女孩。第 2 列安排出生月份,除 2 月 4 孩外,每月 3 孩; 以  $(a_{i2}+2)/3$  的地板函数值显示, 其中  $(37+2)/3=13$ , 转换为 2。其余列考虑少数的特殊情况,以 19 为对称布点; 剩下的空格,用多水平值填补。思路已经在均匀表上用不同灰度显示出。生成的 37 行 10 因素方案安排表如表 12 中所示。

### 3、生成因素水平安排表

将表 12 中每一行转换成一个查询条件,通过关系数据库的 SQL 查询语句,实现寻求满足条件的记录。在有多个记录的情况下,可以选择最符合实际情况的一个记录,或者作更多属性的进一步查询。在没有满足条件的记录存在的情况下,只有改变方案的安排。为了具

	1	2	3	4	5	6	7	8	9	10
1	10	5	4	2	3	9	7	8	1	6
2	20	10	8	4	6	18	14	16	2	12
3	30	15	12	6	9	27	21	24	3	18
4	3	20	16	8	12	36	28	32	4	24
5	13	25	20	10	15	8	35	3	5	30
6	23	30	24	12	18	17	5	11	6	36
7	33	35	28	14	21	26	12	19	7	5
8	6	3	32	16	24	35	19	27	8	11
9	16	8	36	18	27	7	26	35	9	17
10	26	13	3	20	30	16	33	6	10	23
11	36	18	7	22	33	25	3	14	11	29
12	9	23	11	24	36	34	10	22	12	35
13	19	28	15	26	2	6	17	30	13	4
14	29	33	19	28	5	15	24	1	14	10
15	2	1	23	30	8	24	31	9	15	16
16	12	6	27	32	11	33	1	17	16	22
17	22	11	31	34	14	5	8	25	17	28
18	32	16	35	36	17	14	15	33	18	34
19	5	21	2	1	20	23	22	4	19	3
20	15	26	6	3	23	32	29	12	20	9
21	25	31	10	5	26	4	36	20	21	15
22	35	36	14	7	29	13	6	28	22	21
23	8	4	18	9	32	22	13	36	23	27
24	18	9	22	11	35	31	20	7	24	33
25	28	14	26	13	1	3	27	15	25	2
26	1	19	30	15	4	12	34	23	26	8
27	11	24	34	17	7	21	4	31	27	14
28	21	29	1	19	10	30	11	2	28	20
29	31	34	5	21	13	2	18	10	29	26
30	4	2	9	23	16	11	25	18	30	32
31	14	7	13	25	19	20	32	26	31	1
32	24	12	17	27	22	29	2	34	32	7
33	34	17	21	29	25	1	9	5	33	13
34	7	22	25	31	28	10	16	13	34	19
35	17	27	29	33	31	19	23	21	35	25
36	27	32	33	35	34	28	30	29	36	31
37	37	37	37	37	37	37	37	37	37	37

体说明这一点，举出 SQL 查询语句的范例如下：

```

Creat Sql View junyun1i As ;
Select * from baby03810 where xb=1 and yue==2 and yunz=2 and
tiz=2 and yunc>3 and canc=1
Creat Sql View junyun1j As ;
select * from junyun1i where mqnl=1 and (fqnl=3 or fqnl=4) and
mqwh=1 and fqwh=2

```

表 2 出生性别比 10 因素拟水平方案安排表

	孩	月	孕周	体重	母年龄	母文化	父年龄	父文化	孕次	产次
1	男	2	37-41W	2.5-4kg	25岁下	大学以上	31岁以上	中等文化	4次以上	1次
2	男	4	37-41W	2.5-4kg	25岁下	大学以上	26-30岁	未登录	1次	1次
3	女	5	37-41W	2.5-4kg	25岁下	大学以上	26-30岁	未登录	3次	1次
4	男	7	37-41W	4kg以上	25岁下	大学以上	31岁以上	中等文化	1次	1次
5	男	9	37-41W	2.5-4kg	25岁下	中等文化	26-30岁	大学以上	2	1次
6	女	10	37-41W	2.5-4kg	25岁下	中等文化	31岁以上	未登录	1次	1次
7	女	12	37-41W	2.5-4kg	26-30岁	中等文化	31岁以上	大学以上	3次	2次
8	男	1	37-41W	2.5-4kg	26-30岁	中等文化	25岁以下	未登录	1次	1次
9	男	3	37-41W	2.5-4kg	31岁以上	中等文化	31岁以上	大学以上	2	1次
10	女	5	37下41上	2.5-4kg	26-30岁	未登录	31岁以上	中等文化	1次	1次
11	女	6	37-41W	2.5-4kg	26-30岁	未登录	26-30岁	未登录	1次	1次
12	男	8	37-41W	2.5-4kg	26-30岁	未登录	31岁以上	未登录	1次	1次
13	男	10	37-41W	2.5-4kg	26-30岁	未登录	26-30岁	中等文化	1次	1次
14	女	11	37-41W	2.5-4kg	26-30岁	未登录	26-30岁	大学以上	2	1次
15	男	1	37-41W	4kg以上	26-30岁	未登录	31岁以上	未登录	1次	1次
16	男	2	37-41W	2.5-4kg	31岁以上	未登录	26-30岁	大学以上	1次	1次
17	女	4	37-41W	2.5-4kg	26-30岁	未登录	31岁以上	未登录	1次	1次
18	女	6	37-41W	2.5-4kg	26-30岁	未登录	26-30岁	未登录	2	1次
19	男	7	37-41W	2.5kg以下	26-30岁	未登录	26-30岁	中等文化	1次	1次
20	男	9	37-41W	2.5-4kg	26-30岁	未登录	26-30岁	未登录	2	1次
21	女	11	37-41W	2.5-4kg	26-30岁	未登录	26-30岁	大学以上	1次	1次
22	女	12	37-41W	2.5-4kg	26-30岁	未登录	31岁以上	中等文化	1次	1次
23	男	2	37-41W	2.5-4kg	26-30岁	未登录	31岁以上	中等文化	1次	1次
24	男	3	37-41W	2.5-4kg	26-30岁	未登录	25岁以下	未登录	2	2次
25	女	5	37-41W	2.5-4kg	31岁以上	未登录	31岁以上	未登录	1次	1次
26	男	7	37-41W	2.5-4kg	26-30岁	未登录	26-30岁	未登录	1次	1次
27	男	8	37下41上	2.5-4kg	26-30岁	未登录	26-30岁	未登录	1次	1次
28	女	10	37-41W	4kg以上	26-30岁	中等文化	31岁以上	中等文化	1次	1次
29	女	12	37-41W	2.5-4kg	26-30岁	中等文化	25岁以下	中等文化	2	1次
30	男	1	37-41W	2.5-4kg	26-30岁	中等文化	31岁以上	大学以上	1次	1次
31	男	3	37-41W	2.5-4kg	26-30岁	中等文化	31岁以上	未登录	3次	1次

32	女	4	37-41W	2.5-4kg	25岁下	中等文化	26-30岁	中等文化	1次	1次
33	女	6	37-41W	2.5-4kg	25岁下	大学以上	26-30岁	未登录	2	1次
34	男	8	37-41W	2.5-4kg	25岁下	大学以上	26-30岁	未登录	1次	1次
35	男	9	37-41W	2.5-4kg	25岁下	大学以上	26-30岁	大学以上	3次	1次
36	女	11	37-41W	2.5-4kg	25岁下	大学以上	31岁以上	未登录	1次	1次
37	女	2	37-41W	2.5kg以下	31岁以上	大学以上	26-30岁	大学以上	4次以上	1次
	1	2	3	4	5	6	7	8	9	10

意思是从 baby03810 数据库表中查找满足方案表第一行除了父母信息之外的条件，存放于临时表 junyun1i 中，当此表有较多的记录时，进一步根据父母信息进行进一步的查询，结果放在 junyun1j 中。如果没有得到记录，比如第 7 行：孕次大于 3 而产次为第 2 次的生育女孩者，没有出现在 2002 年 12 月份，但是，出现在 2003 年的 4 月份；在第 17 行，4 月和 12 月都有满足的记录；于是，我们就把 7 行和 17 行的月份对换一下。继续做完 37 次查询，每次选择一个记录，集中起来作为最后的样本。变尺度和均匀设计数据库请见<sup>[6] [7]</sup>。

#### 4、样本示意

生成样本的前 20 行记录如表 13 中所示，从中可看出实际运算的数据结构。

表 13 样本的前 20 行记录

	bh	xb	yue	fqn1	yunz1	yunz	tiz	mqn1	mqn1	mqwh	fqn1	fqwh	yunc	canc	xm	tbsj	cszh
1	12739	1	2	2	40	2	2	3	29	0	28	0	4	1	王宇涵	03-2-28	13507175892
2	13366	1	4	2	40	2	2	2	24	0	27	0	1	1	徐弘毅	03-4-29	13871096479
3	13430	2	5	2	41	2	2	4	32	0	26	0	3	1	袁雯静	03-7-23	9420006474
4	10084	1	7	3	39	2	3	2	25	1	31	1	1	1	叶凌峰	02-8-12	88918368
5	10677	1	9	2	40	2	2	2	25	1	26	1	2	1	彭伟恒	02-9-17	88217656
6	11256	2	10	3	39	2	2	2	24	2	33	2	1	1	熊承宇	02-10-29	88044954
7	13230	2	*4	3	39	2	2	3	27	0	34	0	3	2	戴玉婕	03-4-22	88096745
8	12523	1	1	1	38	2	2	3	29	0	25	0	1	1	司安骏	03-3-26	0048389
9	13041	1	3	4	40	2	2	4	35	0	41	0	2	1	许瞳	03-4-24	0014844
10	13390	2	5	2	42	3	2	3	28	0	28	0	1	1	周熙辰	03-7-18	9420006418
11	13445	2	6	2	40	2	2	3	27	0	29	0	1	1	刘姝彤	03-6-20	9420006036
12	10406	1	8	3	38	2	2	3	28	0	35	0	1	1	邱 玢	02-11-21	88850749
13	11081	1	10	2	39	2	2	3	26	1	29	2	1	1	陈天毅	02-10-18	88019367
14	11604	2	11	2	41	2	2	3	30	1	30	1	2	1	叶琬婷	02-11-22	87884703
15	12450	1	1	3	41	2	3	3	29	0	33	0	1	1	杨成才	03-1-29	83227981
16	12777	1	2	3	39	2	2	4	35	1	33	1	1	1	向春岭	03-3-18	420190736
*17	13323	2	4	3	39	2	2	4	32	0	32	0	1	1	王馨竹	03-4-30	88319523
17	11880	2	12	4	39	2	2	3	29	0	39	0	1	1	张奕琳	03-1-2	87713531
18	13613	2	6	2	40	2	2	3	27	0	29	0	2	1	向羽	03-7-12	9420006327
*19	11506	1	11	1	40	2	1	3	27	2	25	2	1	1	何哲曦	02-11-13	88888
20	10919	1	9	2	38	2	2	3	27	2	28	2	2	1	熊海枫	02-10-8	88864257

## 5、基于样本上的对照工作

显然，性别比是不会精确的。笔者打算做决策树的研究，结果将在会议期间报告（或会下交谈）时公布。

## 参考文献

- [1] 杨小莉 2003. 11 武昌区最近一年出生性别比数据分析, 中华医学丛刊( 武汉)
- [2] United Nations. 1955 Method of Appraisal of Quality of Basic Data for Population Estimates, Manual II ST/SOA/Series A /23
- [3] 马赢通. 1993. 5. 24 关于出生性别比与人口性别比的若干问题, 中国人口报
- [4] 马赢通. 冯立天. 1998 再论出生性别比若干问题, 人口与经济( 5)
- [5] 胡良平. 1996. 3 现代统计学与 SAS 应用, 军事医学科学出版社( 北京)
- [6] 方开泰 马长兴 2001 9 正交与均匀试验设计, 科学出版社( 北京)
- [7] 钱 进 2002 8 均匀设计原理及自动实现, 中国科学文化出版社( 香港)