

“当代中国妇女地位研究”抽样调查 方案及抽样效果评估

高嘉陵 陈 萍

【提要】 本文介绍“当代中国妇女地位研究”抽样调查的方案设计,并对方案的精度及设计效果进行了评估。在第一部分详细地叙述了调查所采用的分层四阶不等概率抽样方法及二重抽样技术,讨论了样本量的确定与分配。第二部分详细地讨论了总体目标量的估计及其方差估计的方法,给出了具体的计算公式,最后列出了若干目标量的实际精度和设计效果并就此作了讨论与分析。

【作者】 高嘉陵 中国社会科学院人口研究所副研究员;陈萍 中国社会科学院人口研究所副研究员。

“当代中国妇女地位研究”抽样调查(简称妇女调查)是由中国社会科学院人口研究所承担的联合国人口基金CPR/90/P06项目的研究课题之一,是中国社会科学院的重点项目。该项目旨在对中国当代妇女地位和妇女发展等问题进行实证的和理论的研究。该项目联合上海、山东、广东、陕西、宁夏、吉林、四川、河北、湖北社会科学院和复旦大学、杭州大学的人口研究所11个科研单位于1991年在上述省(市、区)进行了全省或地区性的问卷抽样调查,调查时点为1991年10月1日。

1. 抽样方案

1.1 概述

“妇女调查”是以各省(市、区)的总人口为抽样总体(去除了部分边远人口稀少地区),采用了分层四阶段不等概率抽样方法及二重抽样技术。

在省(市、区)内分城市和农村二个区域(大层)抽样,并将城市和农村作为二个总体进行目标量估计。城市又按人口规模分层,分为大、中、小城市(或大、小城市)三层(或二层),农村按所处地理环境分为三(或二)层。多阶抽样的抽样误差主要来源于第一阶的抽样误差,一般不回放的不等概率抽样其误差较小,因此,在层内第一阶采用不放回的与人口数成比例的不等概率抽样(πPS)方法抽取二个市(县),具体使用的是德宾(Durbin)方法;第二阶在市(县)内采用放回的与人口数成比例的不等概率抽样(PPS)方法抽取若干街道(乡镇);第三阶在街道(乡镇)中同样采用 PPS 方法抽取2个居(村)委会;第四阶在居(村)委会中采用等距抽样(亦称系统抽样)方法抽取样本户。

由于调查对象是20~54岁已婚妇女与丈夫生活在一起的夫妻对(简称合格夫妻对),而符合调查要求的夫妻对事先是没有统计数据的。为了有针对性的进行调查,提高调查效率,我们采用了二重抽样(亦称二相抽样)技术。首先进行第一重样本抽样,估计合格夫妻对在各阶人口和总体人口中的比例。做第一重抽样时,在最后一阶(第四阶)进行对居(村)委

会的整群抽样,在居(村)委会查阅户口登记卡统计出居(村)委会中合格夫妻对的比例,由此推断省(市、区)总体中合格夫妻对的比例。第二重抽样时,在用上述方法进行一、二、三阶抽样后,第四阶在居(村)委会所有合格夫妻对的家庭中抽取实际调查的家庭户。

1.2 样本量的确定与分配

一般抽样调查的样本量是由对估计量所要求的精度并在此精度下以所花费的成本最低而确定的。但实际情况往往是进行调查的资金、人力等客观条件和调查的样本规模已经确定。我们的任务便是在已有的样本规模下,采用各种有效方法提高抽样精度和设计效率。

现确定各省(市、区)的样本量为1500~1400合格夫妻对,城乡各分配750~720夫妻对,为便于自加权处理样本,各市(县)、街道(乡镇)和居(村)委会都按等量分配样本,最后一阶每个居(村)委会抽取25~30合格夫妻对。

1.3 抽样方法与操作

省(市、区)城乡层内第一阶用德宾方法抽取2个市(县)。设层内有 N 个市(县),每个市(县)的人口数为 M_i ($i=1, 2, \dots, N$)。在抽样操作时所使用的参数列表如下:

表1 抽样参数

i	M_i	$z_i = N_i/M$	$Z_i = \sum_{j=1}^i z_j$	$x_i (i \neq K)$	X_i
1	M_1	z_1	$Z_1 = z_1$	$x_1 = z_K \left(\frac{1}{1-2z_1} + \frac{1}{1-2z_K} \right)$	$X_1 = x_1/D$
2	M_2	z_2	$Z_2 = z_1 + z_2$	$x_2 = z_K \left(\frac{1}{1-2z_2} + \frac{1}{1-2z_K} \right)$	$X_2 = (x_1 + x_2)/D$
\vdots	\vdots	\vdots	\vdots	\vdots	\vdots
K	M_K	z_K	$Z_K = \sum_{j=1}^K z_j$		
\vdots	\vdots	\vdots	\vdots	\vdots	$X_L = (x_1 + x_2 \dots x_L)/D$
L	M_L	z_L			\vdots
\vdots	\vdots	\vdots	\vdots	\vdots	\vdots
N	M_N	z_N	$Z_N = \sum_{j=1}^N z_j$	$x_N = z_K \left(\frac{1}{1-2z_N} + \frac{1}{1-2z_K} \right)$	$X_N = \sum_{i \neq K}^N x_i/D$

$$\text{其中 } M = \sum_{i=1}^N M_i, \sum_{i=1}^N z_i = 1 \quad D = 1 + \sum_{i=1}^N \frac{z_i}{1-2z_i} \quad (1)$$

抽样操作时,首先计算 z_i 和 Z_i 列,发生第一个 $[0, 1]$ 中的随机数 α_1 ,若 $Z_{K-1} < \alpha_1 \leq Z_K$,则第 K 个单位被抽中,这是按概率 z_i 抽取的第一个样本单位市(县)。然后计算表内 x_i 列及 D 值与 X_i 列,发生第二个 $[0, 1]$ 中的随机数 α_2 ,若 $X_{L-1} < \alpha_2 \leq X_L$,则第 L 个单位被抽中,这时第二个单位抽中的概率与 x_i 成比例,并且比例除数为 D 。

上述方法中要求 $M_i < \frac{1}{2}M$,为保证这一点,在某些省份一个城市人口大于层内总人口的一半时,将城市分为若干区,以区为抽样单位参加抽样。

用德宾方法抽取二个样本单位时,任一单位 i 第一次被抽中和第二次被抽中的无条件概率均等于 z_i ,因此,单位 i 被抽中(第一次抽中或第二次抽中)的概率 π_i 为

$$\pi_i = 2z_i \quad (2)$$

而样本包含 i, j 两个单位(不计抽中的顺序)的概率 π_{ij} 为

$$\pi_{ij} = \frac{z_i z_j (1 - z_i - z_j)}{D(1 - 2z_i)(1 - 2z_j)} \quad (3)$$

其中 D 见公式(1)。

以上公式(1)、(2)、(3)将在目标量估计时使用。

第二、三阶采用放回的与人口数成比例的不等概率(PPS)方法抽取街道(乡镇)、居(村)委会。

第四阶居(村)委会抽取样本户的等距抽样, PPS 方法与等距抽样方法从略。

2. 目标量的估计及其方差估计

2.1 目标量

本次调查问卷的变量绝大多数是分类变量即定性变量,因此在进行数据汇总时绝大多数指标是总体比例型的目标量,如合格夫妻对占总体的比例、合格夫妻对或妻子、丈夫的某种特征分类在合格夫妻对中所占的比例等。在计算精度和设计效应时,我们只选择丈夫、妻子的年龄、文化分布及合格夫妻对的家庭人均月收入分布指标做分析,以提供对抽样方案和抽样结果做出评估的定量依据。

2.2 记号

为给出目标量的估计方法,首先介绍本文中所使用的记号的意义。

使用英文字母如 y, p, \dots 等表示指标量。大写字母为总体指标量,小写字母为样本指标量,字母上有 \wedge 符号的为估计量。字母下标编号表示某一层次的指标值(或估计值),无下标的字母表示省(市、区)的指标量(或估计值)。例如: y ——样本指标量; \hat{Y} ——总量估计量; P ——样本比例型指标; \hat{P} ——比例估计量; \hat{P}' ——合格夫妻对比例估计量; M ——人口数; L ——户数; n ——样本数; H ——合格夫妻对数。

下标编号有:

无下标——省(市、区)城(乡); h ——层编号; i ——层内市(县)编号; j ——市(县)内街道(乡镇)编号; k ——街道(乡镇)内居(村)委会编号。

如某省 h 层 i 市的人口数表示为 M_{ki} 。在做层内分析时,有时省略 h 编号。

2.3 市(县)的目标量估计及其方差估计

根据抽样方案,市(县)内抽取街道(乡镇)和下一阶抽取居(村)委会均采用放回 PPS 方法。

在进行第一重抽样时,对居(村)委会整群调查合格夫妻对的比例,若居(村)委会的人口规模差别不大,则样本可看作近似自加权的,这样市(县)合格夫妻对的比例估计量为

$$\hat{P}_i = \frac{\sum_k \sum_j P_{ijk}}{K \cdot J} \quad (4)$$

其中 $P_{ijk} = \frac{H_{ijk}}{L_{ijk}}$ 为 i 市 j 街道 k 居委会合格夫妻对总数 H_{ijk} 占居委会总户数 L_{ijk} 的比例,

($i=1, 2, \dots, J, k=1, 2, \dots, K$)

J 为市(县)内样本街道(乡镇)个数;

K 为街道(乡镇)内样本居委会个数;

\hat{P}_i 的方差估计近似为:

$$v(\hat{P}_i) = \frac{\hat{P}_i(1-\hat{P}_i)}{L_i} \quad (5)$$

其中 $L_i = \sum_{j=1}^J \sum_{k=1}^K L_{ijk}$ 为层内市(县)所有样本居(村)委会的户数总和。

在第二重抽样时,我们是在合格夫妻对中抽取样本户,合格夫妻对中分类比例是以合格夫妻对为总体的比例型估计量。

在多阶PPS抽样中只要各阶等样本分配,最后一阶抽样的样本量相等,则可保证最低一级样本单位被抽中的概率相等,从而获得的样本是自加权的。在第二重抽样时,第一、二、三阶都是以人口数为比例的概率抽取样本单位的,而最低一级样本户却是合格夫妻对,如为使每一对合格夫妻被抽中的概率相等,严格地讲,各居(村)委会中的样本户就不应是等样本分配的。我们为了计算和操作方便,不必(或无条件)在第一重样本抽样后计算出合格夫妻对各层次的比例估计,再以各层次合格夫妻对数为比例的概率进行第二重样本抽样。所以在此假设在同一市(县)内居(村)委会中合格夫妻对的比例是基本相同的,这样各居(村)委会等样本抽取合格夫妻对,合格夫妻对样本也就是近似自加权的。因而市(县)内合格夫妻对某种特征分类的比例估计近似为:

$$\hat{P}'_i = \frac{a_i}{n_i} = \frac{\sum_{j=1}^J \sum_{k=1}^K a_{ijk}}{\sum_{j=1}^J \sum_{k=1}^K n_{ijk}} \quad (j=1,2,\dots,J \quad k=1,2,\dots,K) \quad (6)$$

其中 a_{ijk} 为 i 市(县)、 j 街道(乡镇)、 k 居(村)委会中合格夫妻对中某种特征分类的样本数, n_{ijk} 为居(村)委会中的样本夫妻对数。

市(县)内合格夫妻对分类比例估计量的方差估计近似为

$$v(\hat{P}'_i) = \frac{\hat{P}'_i(1-\hat{P}'_i)}{n_i} \quad (7)$$

2.4 省(市、区)层内目标量估计及其方差估计

省(市、区)层内第一阶采用德宾方法抽取二个市(县)。德宾方法可应用霍维茨—汤姆森(Harvitz—Thompson)估计量做目标量的估计。

$$\hat{Y}_h = \frac{y_{h1}}{\pi_{h1}} + \frac{y_{h2}}{\pi_{h2}} \quad (8)$$

$$\text{其中 } \pi_{h1} = 2z_{h1} = \frac{2M_{h1}}{M_h}, \quad \pi_{h2} = 2z_{h2} = \frac{2M_{h2}}{M_h}$$

它的方差估计应用耶茨—格伦迪—森(Yates—Grundy—Sen)的无偏估计量,并应用多阶抽样方差公式,只考虑市(县)间和市县内方差,街道(乡镇)、居(村)委会内的方差相对来说要小得多,可忽视不计。则省(市、区)层内估计量的方差估计应用公式为:

$$v(\hat{Y}_h) = \left(\frac{\pi_{h1}\pi_{h2}}{\pi_{h12}} - 1 \right) \cdot \left(\frac{y_{h1}}{\pi_{h1}} - \frac{y_{h2}}{\pi_{h2}} \right)^2 + \sum_{i=1}^2 \frac{M_{hi}^2(1-f_{hi})s_{hi}^2}{n_{hi}\pi_{hi}} \quad (9)$$

其中 π_{h1} , π_{h2} 见公式(8), π_{h12} 见公式(3); f_{hi} 为市(县)内样本抽样比,当 $f_{hi} \leq 1$ 时可忽略; $s_{hi}^2 = P_{hi}(1-P_{hi})$ 为市(县)内样本方差。

对于第一重样本,省(市、区)合格夫妻对比例估计为

$$\hat{P}_h = \hat{Y}_h / M_h = \frac{1}{2}(\hat{P}_{h1} + \hat{P}_{h2}) \quad (10)$$

它的方差估计为:

$$v(\hat{P}_h) = \frac{v(\hat{Y}_h)}{M_h^2} = A_i (\hat{P}_{h1} - \hat{P}_{h2}) + \sum_{i=1}^2 \frac{M_{hi} \hat{P}_{hi} (1 - \hat{P}_{hi})}{2M_h \cdot n_{hi}} \quad (11)$$

其中 A_i 是用人口数代入公式 (9) 推导出的系数。

$$A_i = \frac{D(M_h - 2M_{h1})(M_h - 2M_{h2})}{4M_h(M_h - M_{h1} - M_{h2})} - \frac{1}{4} \quad (12)$$

其中 D 见公式 (1)。

对于第二重样本, 省(市、区)层内合格夫妻对之中的比例估计量为:

$$\begin{aligned} \hat{P}'_h &= \frac{\hat{Y}_h}{H_h} = \frac{\hat{Y}_h}{M_h \cdot \hat{P}_h} = \frac{1}{2} \left(\frac{M_{h1} \cdot \hat{P}_{h1} \cdot \hat{P}'_{h1}}{M_h \cdot \hat{P}_h \cdot \pi_{h1}} + \frac{M_{h2} \cdot \hat{P}_{h2} \cdot \hat{P}'_{h2}}{M_h \cdot \pi_{h2} \cdot \hat{P}_h} \right) \\ &= \frac{1}{2\hat{P}_h} (\hat{P}_{h1} \cdot \hat{P}'_{h1} + \hat{P}_{h2} \cdot \hat{P}'_{h2}) \end{aligned} \quad (13)$$

它的方差估计为:

$$\begin{aligned} v(\hat{P}'_h) &= \frac{v(\hat{Y}'_h)}{M_h^2 \cdot \hat{P}_h^2} = \frac{A_i}{P_h^2} (\hat{P}_{h1} \cdot \hat{P}'_{h1} - \hat{P}_{h2} \cdot \hat{P}'_{h2}) \\ &\quad + \sum_{i=1}^2 \frac{M_{hi} \cdot \hat{P}_{hi}^2 \cdot \hat{P}'_{hi} (1 - \hat{P}'_{hi})}{2M_h \cdot \hat{P}_h^2 \cdot n_{hi}} \end{aligned} \quad (14)$$

其中 A_i 见公式 (12)。

从公式 (11) 和 (14) 可以看到层内方差的主要来源于第一阶市(县)间的方差, 即公式中的第一项, 而第二阶市(县)内的方差均为以市(县)为总体时方差的 π_i 倍 (π_i 是小于 1 的数)。由此可知第二、三阶中样本单位内的方差成分将更小, 所以可忽略不计。从公式 (13)、(14) 中还可以看到, 省(市、区)层内合格夫妻对的指标估计量还与层内和二个市(县)的合格夫妻对比例有关, 如果 \hat{P}_h , \hat{P}_{h1} , \hat{P}_{h2} 近似相等, 则公式 (13)、(14) 近似等于公式 (10)、(11)。为简化计算, 省(市、区)层内合格夫妻对之中的比例估计量使用的是公式 (10), 它的方差估计使用公式 (14)。

2.5 省(市、区)目标量的估计及方差估计

得到省(市、区)城乡各层的估计后, 省(市、区)城乡的估计即可用分层抽样公式计算。

$$\hat{P} = \sum_{h=1}^g W_h \hat{P}_h \quad (h=1, 2, \dots, g) \quad (15)$$

$$v(\hat{P}) = \sum_{h=1}^g W_h^2 v(\hat{P}_h) \quad (h=1, 2, \dots, g) \quad (16)$$

其中 W_h 为层权, 对于第一重抽样, $W_h = M_h/M$ 为人口数权重, 第二重抽样的层权 $W_h = H_h/H$ 为合格夫妻对的权重, g 为层数。

现将上海、陕西、山东、广东、宁夏、吉林 6 个省(市、区)做为推论总体, 将各省(市、区)的估计用以上分层抽样公式加权求和, 权数为各省合格夫妻对数与 6 省(市、区)合格夫妻对数之比, 即计算出 6 省(市、区)调查的目标量估计与方差估计, 从而可进一步计算总的调查指标的精度和设计效应。

3. 目标量的精度与设计效应

本调查的调查方式是研究人员和受培训的调查员入户面访,因此问卷的回收率达100%,有效问卷99%以上,并且问卷中绝大多数问题的回答率也在95%以上,所以在数据汇总和计算精度时将不考虑不回答因素的影响。

本调查是分层四阶不等概率的复杂抽样调查,不能直接使用简单随机抽样的精度公式,调查指标的实际精度必须按照上述所给出的公式,根据实际调查数据计算出方差而得。

在获得了各省(市、区)和6省(市、区)的目标量估计的方差估计后,通过计算标准差:

$$s(\hat{P}) = \sqrt{v(\hat{P})} \quad (17)$$

求出在95%置信水平下的相对误差

$$r(\hat{P}) = \frac{u_{\alpha} s(\hat{P})}{\hat{P}} \quad (18)$$

其中 u_{α} 是标准正态分布的 α 双侧分位数,在95%置信水平下, $u_{\alpha}=1.96$,则估计量的精度为:

$$A = 1 - r(\hat{P}) = 1 - \frac{u_{\alpha} s(\hat{P})}{\hat{P}} \quad (19)$$

设计效应 $deff$ 是另一个用来衡量复杂抽样的设计效率的量。它是复杂抽样的实际方差与在相同样本下采用简单随机抽样时方差的比值。它反应了在复杂抽样时设计方法不同引起的样本量的效率的不同,而样本效率直接与调查成本相关。因此设计效应是评价设计方案优劣的重要指标之一。

对于比例型估计量 \hat{P} ,简单随机抽样的方差估计为 $v(\hat{P}) = \hat{P}(1-\hat{P})/n$,于是设计效应的估计为:

$$deff = \frac{nv(\hat{P})}{\hat{P}(1-\hat{P})} \quad (20)$$

其中 n 为样本量, \hat{P} 为比例估计量, $v(\hat{P})$ 为比例估计的方差估计。

经过以上公式计算,6个省(市、区)的合格夫妻对比例以及6省(市、区)总体夫妻年龄、文化分布和家庭人均月收入分布的精度和设计效应列表如下:

表2 6省(市、区)合格夫妻对占总人口的比例的精度和设计效应

地区	分类	总体	上海	陕西	山东	广东	宁夏	吉林
城市	比例%	18.98	15.76	12.19	22.34	16.35	23.73	21.10
	精度%	96.95	90.65	92.31	95.88	91.06	92.39	94.42
	设计效应	4.41	3.40	2.55	2.21	6.41	7.78	1.98
农村	比例%	19.58	18.98	18.39	22.09	17.48	19.05	23.33
	精度%	96.85	92.92	88.64	97.55	92.41	93.00	91.50
	设计效应	3.95	2.21	6.26	0.34	4.40	4.94	5.64

注:本文中所指的精度均在95%置信水平下计算的结果。

表3 6省(市、区)合格夫妻对夫妻年龄分布及其精度和设计效应

地区	分类	20~24	25~29	30~34	35~39	40~44	45~49	50~54	55+
城市 (丈夫)	比例%	1.51	14.30	19.07	23.16	15.43	11.84	8.73	5.96
	精度%	65.68	85.92	93.43	92.72	90.26	84.24	85.50	82.70
	设计效应	2.12	3.88	1.19	1.87	2.03	3.19	2.36	2.22
城市 (妻子)	比例%	4.21	17.02	20.83	23.10	15.97	11.05	7.82	
	精度%	74.88	91.95	92.49	95.41	85.76	85.04	85.62	
	设计效应	3.25	1.56	1.74	0.74	4.52	3.26	1.91	
农村 (丈夫)	比例%	5.47	19.18	17.56	20.54	16.90	10.71	6.12	3.52
	精度%	73.45	91.73	90.56	93.07	88.47	86.65	71.92	79.90
	设计效应	4.80	1.91	2.23	1.46	3.18	2.51	6.05	1.73
农村 (妻子)	比例%	8.12	20.34	19.50	20.58	16.62	9.72	5.12	
	精度%	86.95	90.72	91.60	87.52	90.57	80.25	73.16	
	设计效应	1.77	2.59	2.01	4.75	2.09	4.94	4.18	

表4 6省(市、区)合格夫妻对夫妻文化程度分布及其精度和设计效应

地区	分类	文盲	识字不多	小学	初中	高中	中专	大专	大专以上
城市 (丈夫)	比例%	0.30	0.28	4.73	29.23	22.36	15.21	18.34	9.55
	精度%	7.87	42.52	64.70	83.99	91.57	80.32	85.08	71.16
	设计效应	3.00	1.09	7.26	12.41	2.40	8.15	5.86	10.33
城市 (妻子)	比例%	0.71	1.35	8.32	33.25	26.14	16.16	10.26	3.81
	精度%	70.42	66.74	73.31	86.34	91.78	84.38	74.85	57.04
	设计效应	0.73	1.78	7.58	10.91	2.81	5.51	8.48	8.57
农村 (丈夫)	比例%	3.98	4.80	32.62	42.19	14.60	1.37	0.35	0.09
	精度%	54.23	70.58	87.77	91.68	86.49	72.91	15.06	—
	设计效应	10.22	5.14	8.52	5.94	3.67	1.20	2.98	3.19
农村 (妻子)	比例%	24.00	12.57	27.90	20.61	7.55	0.41	0.83	0.21
	精度%	73.87	82.46	89.26	85.54	75.89	—	—	—
	设计效应	25.38	5.21	5.25	8.92	5.59	5.99	31.35	3.27

注：“—”为精度极差。

表5 6省(市、区)合格夫妻对家庭人均月收入分布及其精度和设计效应

地区	分类	≤50	51~100	101~200	201~300	301~400	>400
城市	比例%	3.17	29.79	37.23	18.37	10.56	0.87
	精度%	71.06	88.28	91.25	88.50	49.34	—
	设计效应	3.22	6.83	5.33	3.49	35.54	114.22
农村	比例%	44.72	33.95	13.26	4.72	2.91	0.43
	精度%	78.46	80.27	77.24	75.18	80.83	47.37
	设计效应	44.18	23.54	9.32	3.59	1.30	1.41

注：“—”为精度极差。

精度和设计效应这两个评价抽样质量的指标都与样本量、设计方法和样本方差（即指标量实际方差的无偏估计）这三个因素有关。在设计方法相同的情况下，精度这个指标对于样本量较敏感，而设计效应对最高一阶样本单位间的方差较敏感。也就是说，在同一个抽样方案下，某种特征分类的比例估计高，其相应的精度就高，指标精度不满足要求可增加样本量，或将低精度的指标分类与相近特征的分类合并加以解决。如设计效应不满足要求，这主要是样本单位间的差别大所致。由于样本单位间的差别是客观情况的反映，要减少设计效应就要重新选择设计方法，或运用分层分阶的技巧改进样本构成和组织形式。

表2的数据表明，推论总体6省（市、区）合格夫妻对比例的精度城市为96.95%，农村为96.85%，设计效应城市为4.41，农村为3.95。各省（市、区）合格夫妻对比例的精度城乡均超过90%。设计效应在2~7之间，各省（市、区）有所差别。城市中广东、宁夏的设计效应较高，这是由于对居委会做整群抽样，广东、宁夏的居委会人口数较多，样本量大，使得相应简单随机的方差相对小而引起的。农村中陕西、吉林的设计效应较高，山东较低，经从数据中查明这是样本单位间方差的大小引起的。

表3、4、5的数据表明，总体合格夫妻对的指标分布的精度和设计效应在各分组指标中是不同的。为便于在同一样本量水平下进行比较，我们以百分比频率等于10%、设计效应等于3为准来比较各指标分组的情况。表3数据城乡夫妻年龄分布分组指标值在10%以上的精度在85~93%之间，分组指标中有63%的设计效应小于3；表4数据城乡夫妻文化程度分布分组指标值在10%以上的精度在75~90%之间，分组指标中有22%的设计效应小于3，表5数据城乡家庭人均月收入分布分组指标值在10%以上的精度在44~90%，只有两个分组指标的设计效应小于3。这说明，本调查的年龄分布抽样效果最好，文化程度次之，家庭人均月收入的抽样效果最差。这是因为在我们的抽样方法中，样本是按人口规模分层，并按人口数比例的概率抽取的，样本单位之间年龄分布的差别较小，设计效应就小，而其它指标，如家庭人均月收入的差别就可能大，设计效应就增大。如果我们改变抽样方法，如以收入分层，并按收入多少成比例的概率抽样，那么必然能改进收入指标分布的抽样效果。这就要由调查的研究目的和要求来决定。

“妇女调查”是一个综合性的指标较为广泛的抽样调查，在样本量不能十分充分的情况下，不能使每一个指标的抽样效果都令人满意，我们给出以上三种指标分布的精度和设计效应，即给出本调查中指标抽样效果较佳，一般、较差的范围。

参 考 文 献

- 1 [美] William G Cochran, 张尧庭, 吴辉译. 抽样技术. 中国统计出版社
- 2 L. Kish, Survey Sampling, John Wiley & Sons (1965)
- 3 Feng Shi-yong & Wang Si-ping, The design and Data Processing of the Sampling Survey of Children's Situation of China, 1987. Acta Mathematica Applicatae Sinica, Vol 6 (4), 351—360 (1990)
- 4 冯士雍, 杨若勇. 北京地区专业技术人员现状抽样调查的抽样设计, 数据处理方法和精度分析. 应用概率统计. 1991. 7. 4. 425~432
- 5 黄良文, 吴国培. 应用抽样方法. 中国统计出版社, 1991
- 6 尹林生等. 人口抽样调查. 陕西人民出版社, 1991

（本文责任编辑：朱 犁）