World Values Survey, China 2012

Sampling Design

by

Shen Mingming and Yan Jie

Research Center for Contemporary China

Peking University

October 2012

Nov. 5, 2012

The Project

As a participating country team of the World Values Survey, the Research Center for Contemporary China (RCCC) at Peking University implemented the China survey in 2012. The general purpose of the China survey is: through face-to-face questionnaire interviews in a national sample of China containing 40 primary sample units to collect individual level representative data of, from a political cultural perspective, the values and attitudes currently held by Chinese citizens.

Population

The target population covers adults between the ages of 18 and 75, who reside in all 31 provinces of the Chinese Mainland (Hong Kong, Macao and Taiwan are not included).

Sample Size

The sampling units are divided into 40 PSUs, 80 SSUs.
The sample size is determined to be approximately 3,840 eligible individuals, who will be drawn out of the above defined target population in all provinces of China, with at least 2,500 completed, effective samples.

Sampling Method

To meet the requirement of overall coverage of Chinese adults including migrant population, "GPS/GIS Assistant Area Sampling" [1] is used in this survey. Respondents will be sampled through stratified, multi-stage PPS (probability proportional to size) sampling.

Stratification

For the purpose of allocating PSUs across all large regions with different levels of economic development, stratification based on 7 official division of regions in China（Northeast, North, East, Central, South, Northwest, and Southwest）is taken as the first step of the sampling process.
In addition, we have taken into consider the disparity between rural and urban areas in China's social development. In order to lower the disparity between sampling from urban and rural areas in order to fulfill the research goal, this project will do a second stratification based on urban and rural

---

[1] "GPS/GIS Assistant Area Sampling" was developed by the Research Center of Contemporary China at Peking University with Professor Pierre Landry at Yale University, which is particularly advantaged at correcting the bias caused by coverage errors in list-based samples. The bias is especially acute where the level of internal migration is high and where record keeping on the population is not reliable. GPS sampling overcomes the inability to reach migrants in traditional area samples based on household lists. For more details, please refer to Landry and Shen, "Reaching Migrants in Survey Research.", *Political Analysis* (2005) 13:1-22.

characteristics within the first strata.

Therefore, there are 14 layers in total. In order to obtain a self-weighted sample, number of primary sampling units (PSUs) within each stratum is proportional to the population size of that stratum.

Table 1 PSUs allocated in each strata

	2010 population	proportion	PSUs that should be allocated according to the size proportion	PSUs allocated finally	Total PSUs in each st rata	fpc_psu
11 North_urban	77459340	0.05644	2.2576	2	151	0.01325
13 North_rural	85540000	0.06233	2.4932	3	273	0.01099
21 Northeast_urban	71732572	0.05227	2.0907	2	195	0.01026
23 Northeast_rural	35220000	0.02566	1.0265	2	93	0.02151
31 East_urban	211994056	0.15447	6.1788	6	348	0.01724
33 East_rural	175191152	0.12765	5.1062	4	293	0.01365
41 Central_urban	104104200	0.07586	3.0342	3	184	0.01630
43 Central_rural	139220000	0.10144	4.0577	4	200	0.02000
51 South_urban	88901834	0.06478	2.5912	3	131	0.02290
53 South_rural	65260000	0.04755	1.9021	2	122	0.01639
61 Southwest_urban	73679088	0.05369	2.1475	2	120	0.01667
63 Southwest_rural	141940000	0.10343	4.1370	3	389	0.00771
71 Northwest_urban	38798458	0.02827	1.1308	2	95	0.02105
73 Northwest_rural	63350000	0.04616	1.8464	2	261	0.00766
Total	1372390700	1	40	40	2855	0.01401

Sampling Units

Primary Sampling Units (PSUs):

County level administrative units (municipal districts, county-level cities, counties)

Secondary Sampling Units (SSUs):

Half-square minutes (HSM) of latitude and longitude

Tertiary Sampling Units (TSUs):

Spatial square seconds (SSS), approximately 90m*90m

Basic Sampling Units:

Dwellings in the sampled units

Sampling Frames

The sampling frame employed by the primary sampling unit will be taken from the name list of all county-level administrative units and population statistics taken from the <National sub-county Population Statistics 2010>

(published by the Ministry of Public Security, November, 2011 by Qunzhong Publishing House in Beijing).

A GIS dataset will be established as the sampling frame for this project, which will be based on 1) county level population data from the 2010 Census,[2] 2) the most recent and detailed (paper and electronic) maps, 3) the highest possible resolution images from Google Earth. Based on the abive information, the population density is then calculated for each of the HSMs in county level units.

Sampling Processes

1) Out of 2,855 counties in China, 40 counties will be chosen by stratified PPS.

2) Two HSMs will be selected by PPS within each of the selected county.

The measures of size (HSM) used at these stages are the density of the population per sampling unit. We use the points of light data to estimate the population per HSM.

3) Within each of the selected HSM，the number of SSSs （90m*90m）is calculated based on the population density, and then selected the SSSs simple randomly.

4) Trained surveyors equipped with GPS receivers are then sent to locate and enumerate the sampled "spatial square seconds" (SSS). For maintaining equal probabilities of selection across households, all dwellings enumerated in the SSSs will be included in the sample. Using system sampling, we will draw 48 dwellings in each HSM.

5) Respondents will be selected from dwellings using the Kish Grid method[3].

Weighting and Post Stratification

Weighting and post stratification will be done on three major demographic variables, strata, age and gender, as following.

First, calculate probability of every eligible adult to be selected into the sample in each PSU;

---

[2] Although data from the 2000 Census seemed out of date, but it must be used because that it is the only available national dataset that has migrants information throughout province, city, county and township level units.

[3] Leslie Kish, *Survey Sampling* (New York: John Wiley & Sons. 1965), Pp.398-399.

Second, calculate probability of respondents in each of the PSUs;
Third, calculate the weight of each of the respondents;

The final weight is obtained from the post-stratification of respondents in terms of strata, age and gender based on the 2010 Census data.

We will also calculate fpc of the PSU.