**2010 China Survey Technical Report**

1. **Location**

The survey area included 25 provinces in China (except Hong Kong, Macao, Taiwan, Xinjiang, Tibet, Qinghai, Inner Mongolia, Ningxia, and Hainan), covering 94.87% of the Chinese population.

1. **Timetable**

Project: April 1, 2011 – April 20, 2012

Fieldwork: July 1, 2011 – October 31, 2011

1. **Respondents**

Citizens in mainland China over the age of 18, survey carried out through face-to-face interviews.

Double sampling at county and village/neighborhood levels using the China Family Panel Studies (CFPS), Institute of Social Science Survey, Peking University framework. The CFPS sampling frame acts as the household survey sampling frame, which is used to sample a certain number of households. Finally, in each sample household, an individual that meets the survey requirements is randomly selected.

1. **SAMPLING PROCESS AND METHODOLOGY**

(1) Expected Sample Size

The survey intends to achieve an effective sample size of 3500. Samples are selected from the six sampling frames from the CFPS according to proportion of the population. The final effective sample size was 3510, distributed as shown in Table 1.

**Table 1 Effective Sample Distribution**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Level** | **Province or Equivalent Administrative Region** | **2007 Population Statistics** | **Proportion** | **Effective Sample Distribution** |
| 1 | Shanghai | 1858 | 1.51 | 53 |
| 2 | Liaoning | 4298 | 3.49 | 123 |
| 3 | Henan | 9360 | 7.6 | 267 |
| 4 | Guangdong | 9449 | 7.67 | 269 |
| 5 | Gansu | 2617 | 2.13 | 75 |
| 6 | Beijing | 1633 | 77.6 | 2723 |
| Tianjin | 1115 |
| Hebei | 6943 |
| Shanxi | 3393 |
| Jilin | 2730 |
| Heilongjiang | 3824 |
| Jiangsu | 7625 |
| Zhejiang | 5060 |
| Anhui | 6118 |
| Fujian | 3581 |
| Jiangxi | 4368 |
| Shandong | 9367 |
| Hubei | 5699 |
| Hunan | 6355 |
| Guangxi | 4768 |
| Chongqing | 2816 |
| Sichuan | 8127 |
| Guizhou | 3762 |
| Yunnan | 4514 |
| Shaanxi | 3748 |
| **Total** |  | **123128** | **100** | **3510** |

The sampling design uses the CFPS county and village/neighborhood double sampling framework. The entire sampling design uses multi-stage sampling with unequal probabilities. Given 3510 successful samples, at a 95% confidence level, the sampling error is within 1.58%. The distribution of samples for each level and stage is shown in Table 2:

**Table 2 Number of Effective Samples at Each Stage**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | **Sample size** | **First stage** | **Second stage** | **Third stage** | **Fourth stage** |
| Province or equivalent Administrative Region | Expected sample distribution | Sampled counties | Sampled neighborhood or village committees | Sampled households | Sampled individuals |
| Shanghai | 53 | Directly sampled village committee/urban neighborhood committee | Neighborhood or village committee (106) | 15 effective samples of each neighborhood or village committee | 1 individual sampled for each household |
| Liaoning | 122 | 4 |
| Henan | 266 | 6 |
| Guangdong | 268 | 6 |
| Gansu | 75 | 2 |
| Beijing | 2716 | 70 |
| Tianjin |
| Hebei |
| Shanxi |
| Jilin |
| Heilongjiang |
| Jiangsu |
| Zhejiang |
| Anhui |
| Fujian | Neighborhood or village committee (126) |
| Jiangxi |
| Shandong |
| Hubei |
| Hunan |
| Guangxi |
| Chongqing |
| Sichuan |
| Guizhou |
| Yunnan |
| Shaanxi |
| **Effective number of samples** | **3510** | **88** | **234** | **3510** | **3510** |

(2) The Sampling Process

The survey was conducted using double sampling based on the CFPS sampling design,

The multi-stage sampling design is divided into two types. First, Shanghai is a special case, using a three-stage systematic sampling design. Second, the 24 provinces and cities (four large provinces - Guangdong Province, Liaoning Province, Henan Province, Gansu Province - and 20 small provinces) use a four-stage systematic sampling design.

2.1 First Stage

2.1.1 First Stage for Shanghai

Due to the specific characteristics of Shanghai, we did not directly sample counties. According to the CFPS sampling framework, four neighborhoods/villages were sampled using systematic sampling. The village/neighborhood sampling frame is as follows:

i. Each village/neighborhood is divided into urban and rural levels according to coding by the National Bureau of Statistics.

ii. For each of the urban and rural levels, neighborhoods are ranked in descending order on the basis of the GDP for each county to which the village / neighborhood belongs (where there is no GDP data, the proportion of non-agricultural population is used; where there is no GDP or proportion of non-agricultural population data available, population density is used). Next, samples are ranked in order from subdistrict, to county-level city level and county. In each of the subdistrict, county-level city, or county categories, each unit is ranked on the basis of the GDP per capita (where there is no GDP data, the proportion of non-agricultural population is used; where there is no GDP or proportion of non-agricultural population data available, population density is used) in descending order.

2.1.2 First Stage for Guangdong Province, Liaoning Province, Henan Province, and Gansu Province

The first phase of sampling in the four provinces of Guangdong, Liaoning Province, Henan Province, Gansu Province was carried out according to the CFPS sampling frame. Units are first ranked, and then a corresponding number of county samples are drawn from a random starting point using equidistant sampling.

The county sampling frame is as follows:

i. For provincial capitals and other cities, divided into the district level, county-level city and county level, and ranked in descending order from district, county-level city level and county.

ii. For districts, county-level cities and counties, ranked on the basis of the GDP for the prefectural level city to which the district, county-level city/county belongs (where there is no GDP data, the proportion of non-agricultural population is used; where there is no GDP or proportion of non-agricultural population data available, population density is used) in descending order.

iii. Finally, ranked on the basis of the GDP per-capita for each district, county and county-level city (where there is no GDP data, the proportion of non-agricultural population is used; where there is no GDP or proportion of non-agricultural population data available, population density is used) in descending order.

2.1.3 First Stage for 20 Small Provinces

As the sample was taken using the CFPS county sample, 70 county samples were taken using random sampling from the 80 CFPS county samples.

2.2 Second Stage (excluding Shanghai)

In the first stage county/neighborhood and village samples, each village/neighborhood committee is divided into urban and rural levels according to coding by the National Bureau of Statistics. For each of the urban and rural levels, random sampling is used to select 106 neighborhood committees and 128 village committees.

2.3 Household Sampling

For each sampled village and neighborhood, the names and addresses of residents were obtained. The dwelling unit code was matched to the list of residents to create the sampling frame. The household sampling framework was based on the CFPS sampling frame, with a further extended sampling method on the basis of response rates and the situation on the ground to ensure that the target number of samples is taken in each neighborhood or village. Equidistant sampling was used to sample the required number of samples.

Please note: The quality of the final CFPS samples has declined over time. The extended sample used in the survey is adjusted according to response rates in the CFPS survey. The effective urban response rate[[1]](#footnote-1) increased 10% from CFPS questionnaire, while the rural response rate increased by 5%. Taking into consideration funding issues and other practical consideration, the lowest response rate for neighborhoods and villages in urban areas is 50%, compared to 55% in rural areas. Adjustments were made according to the actual conditions found in villages and neighborhoods to ensure that in each unit 15 households were samples.

2.4 Individual Sampling

In each sampled household, KISH random sampling was used to select an individual that met the requirements of the survey.

(3) Sample Size

On the basis of the above,[[2]](#footnote-2) 234 villages and neighborhoods were sampled, giving a total of 5308 samples that were contacted and 3510 effective samples. Using the CSES response rate, the response rate for urban areas was 62%, requiring 2725 contacts. The number of effective samples was 1590, giving an average response rate of 59%. The response rate for rural areas was 77%, requiring 2583 contacts. The number of effective samples was 1920, giving an average response rate of 74%.The average response rate for neighborhoods and villages was 70%, the overall response rate was 66%, and the urban to rural ratio was 45:55.

1. **EXAMINATION OF REPRESENTATIVENESS OF SAMPLE SET AND WEIGHTING**

After data cleaning, the number of successful samples is **3473.** Is the demographic structure of successful samples consistent with that of the entire population? Characteristics of the sampled population are listed below to examine whether or not our sample is representative. Population characteristics considered were gender, age, urban/rural, and region. Figures for the entire population come from China Statistical Yearbook 2011(2012.02.24).

**Sample Representativeness: Gender (before weighting)**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Sample | | Population | Result |
|  | Frequency | Percent | Percent |
| Male | 1824 | 52.5 | 49.8 | Chi square=0.292  P>.05  Consistent with the population |
| Female | 1647 | 47.5 | 50.2 |
| Total | 3471 | 100.0 | 100.0 |

**Sample Representativeness: Age (before weighting)**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Sample | | Population | Result |
|  | Frequency | Percent | Percent |
| 20 – 29 | 684 | 19.7 | 21.6 | Chi square=1.469  P<.05  Not consistent with the population |
| 30 – 39 | 570 | 16.4 | 20.1 |
| 40 – 49 | 862 | 24.8 | 22.2 |
| 50 – 59 | 641 | 18.5 | 17.8 |
| 60 and above | 715 | 20.6 | 18.3 |
| Total | 3472 | 100.0 | 100.0 |

**Sample Representativeness: Urban-rural (before weighting)**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Sample | | Population | Result |
|  | Frequency | Percent | Percent |
| Urban | 1607 | 46.3 | 48.1 | Chi square=0.130  P>.05  Consistent with the population |
| Rural | 1866 | 53.7 | 51.9 |
| Total | 3473 | 100.0 | 100.0 |

**Sample Representativeness: Areas (before weighting)**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Sample | | Population | Result |
|  | Frequency | Percent | Percent |
| Shanghai | 60 | 1.7 | 1.5 | Chi square=1.375  P<.05  Not consistent with the population |
| Liaoning | 117 | 3.4 | 3.5 |
| Henan | 189 | 5.4 | 7.6 |
| Guangdong | 225 | 6.5 | 7.7 |
| Gansu | 45 | 1.3 | 2.1 |
| Other provinces | 2837 | 81.7 | 77.6 |
| Total | 3473 | 100.0 | 100.0 |

As the tables above show, the structures of gender and urban/rural in our sample set are consistent with the entire population. Both age and areas structures in the sample set are not consistent with the entire population. In order to achieve consistency between the sample set and the entire population, we used the raking method to weight our samples. We then verified that the structure of the sample set after weighting was consistent with the entire population.

**Sample Representativeness: Gender (After weighting)**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Sample | | Population | Result |
|  | Frequency | Percent | Percent |
| Male | 1697 | 48.9 | 49.8 | Chi square=0.032  P>.05  Consistent with the population |
| Female | 1774 | 51.1 | 50.2 |
| Total | 3471 | 100.0 | 100.0 |

**Sample Representativeness: Age (After weighting)**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Sample | | Population | Result |
|  | Frequency | Percent | Percent |
| 20 – 29 | 759 | 21.9 | 21.6 | Chi square=0.015  P>.05  Consistent with the population |
| 30 – 39 | 707 | 20.4 | 20.1 |
| 40 – 49 | 768 | 22.1 | 22.2 |
| 50 – 59 | 614 | 17.7 | 17.8 |
| 60 and above | 625 | 18.0 | 18.3 |
| Total | 3472 | 100.0 | 100.0 |

**Sample Representativeness: Urban-rural (After weighting)**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Sample | | Population | Result |
|  | Frequency | Percent | Percent |
| Urban | 1665 | 47.9 | 48.1 | Chi square=0.002  P>.05  Consistent with the population |
| Rural | 1808 | 52.1 | 51.9 |
| Total | 3473 | 100.0 | 100.0 |

**Sample Representativeness: Areas (After weighting)**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Sample | | Population | Result |
|  | Frequency | Percent | Percent |
| Shanghai | 52 | 1.5 | 1.5 | Chi square=0.000  P>.05  Not consistent with the population |
| Liaoning | 122 | 3.5 | 3.5 |
| Henan | 263 | 7.6 | 7.6 |
| Guangdong | 268 | 7.7 | 7.7 |
| Gansu | 73 | 2.1 | 2.1 |
| Other provinces | 2696 | 77.6 | 77.6 |
| Total | 3473 | 100.0 | 100.0 |

1. Here, the effective response rate is a total ratio is equal to – including the empty the households rate - wrong address rate - the refusal rate - failure to contact rate – and rate of households and individuals failing to meet survey conditions. [↑](#footnote-ref-1)
2. The sample is subject to adjustment according to the actual situation [↑](#footnote-ref-2)