# The Application of Dural-System Estimation in Population Census

Zhou Weizhong

*Guangxi survey office of the national bureau of statistics*

*No.2 SiXian RD*

*NanNing(530022),China*

*E-mail:zdsyc@gx.stats.cn*

Population census is a major national conditions and strength investigation, which widely applied by various countries.The census always unavoidably exists omission and misstatement phenomenon, which caused a bad influence on data quality. Therefore, most countries in the world will organize a post enumeration survey after the population census in order to correct census data, increase the data quality of population census; this is the Dural-system evaluation .Currently, the countries which using dural system evaluation roughly follow the same theoretical model, but are different in the concrete operation

Dural-system evaluation has great significance for improving the data quality, therefore, almost all countries in the world will elaborately organize the post enumeration survey after population census in spite of great cost human and material resources. For example, the United States organized one post enumeration survey since the first census in 1790. In 2010, the quantity of the census district which adopted sampling survey reached 300,000, sampling 2.5‰.Thus they treat this survey very important

In 2000 census Dural-system evaluation was attempted in China, 2010 still adopted it, and made improvements, for example, the stratified sign of post enumeration survey is the urban and rural in 2000, In 2010, it had been changed to the mobile population types, this is a progress. But by the objective conditions, it needs to be further improved in a certain way from sampling survey of the view for the theory. One reason is the sample size too few, it seems to fail to take into account the influence of the population scale when considering the mobile population types.It just extracted 402 census district in the whole country, each village about 100 households on average, sampling 0.09 ‰. For example, G province belongs to outflows net populations provinces, has 10 census district, B city belongs to inflows net population distribution, 20 district. But judging from the total population, B city has less than 2 million including floating population. G，also province is 50 million, S provinces has nearly nine million, only got 15 distribution areas. The second reason is because of the little sample, it inevitably cause difficult to post-stratification. One of the theoretical prerequisites for using Dural system evaluation nationwide true population is everyone in the two surveys have the same probability of being registered into the sample .The only means to satisfy population equiprobability sample input is to adopt sample post-stratification, and no doubt the lack of sample size brings obstacles to post-stratification.If directly estimate national population by using

double system without post-stratification, it is not only lacks theoretical basis, but also affected data quality .The third reason is Dural -system evaluation cannot calculate the population components separately. According to the above mentioned the theoretical basis of the second point (i.e. via post-stratification makes two investigations into population equiprobability sample input), with Dural -system evaluation appraise the true number of population, it need to construct Dural -system evaluation only in the internal of each post-stratification, estimate the population of each one, it will get the total population of various provinces or the whole country，then add the total population of each post-stratification。 If do not do as the above mentioned, it will break the request of the sampling estimation theory

During the process of the post enumeration survey of 2000 census, because of the above shortcomings, the Dural- system evaluation had not fully estimated. For the 2010 or 2020 census, how to use Dural- system evaluation estimate the true nationwide population, the author puts forward the following five claims with you, Hope every expert ,scholar and colleague kindly check and advise your comments.

**Firstly Formulate a reasonable sample size scheme**

How many samples should be abstracted of the post enumeration survey in census? how many samples are"reasonable"? Though there exist no absolute standard, but there is always a roughly acceptable range. The author thinks that, not only to understand the national census, the population of each provinces is also an important statistics, therefore, it should treat the province as a unit when design sample size scheme ,so it can make big enough sample size for provinces, so as to undertake post-stratification. It extracted 402 census district in the whole country in 2010 census ,about 40000 housholds,12 millions, sampling 0.09 ‰. If want to achieve the purpose of précising the provinces so as to facilitate post-stratification, with reference to the experience of some other countries in international such as the United States (sample than 2.5‰), sample size should be expanded 20-30 times based on it, make sampling ratio reach 2-3 ‰ or so, each province can assign to 150-250 census district. But Dural- system evaluation inevitable have a defect-great cost of manpower and the physical resource .because of large scale enlarge of sample size

**Secondly Selecting proper signs to carry out post-stratification**

One of premise by using Dural- system evaluation to estimate true population is make two investigations into population sample input equiprobability .The second one is these two survey should be independent, how to ensure the second theoretical premise to be exist, the author will discuss in the later part. For the first theoretical premise (or theoretical assumptions), it is impossible to realize in field sampling.The first reason is the mobile of population, secondly is individual behavior and the design flaws of census form, Moreover, whether it census or post enumeration survey, both in the goal of obtains the true population.Under this goal,it cannot make the name

list in the unit of the person before survey (sampling frame), and will not be able to prior stratification for the total population (take the person as the unit )

However, this does not mean do nothing. If take the total population as the unit and divide into several stratifications in accordance with a certain method , the same probability registered people registered in the same stratification, the different probability registered people into different stratification, and then, estimate "true" population formula respectively in different one of the country, so the probability requirements can be met. It is not difficult to prove that such sample of each stratification can be considered as the sample which once being abstracted from the matching stratification of the overall sampling independently

After understanding why should undertake post-stratification, the next question is what signs should apply. Even if view other countries in the world, In 2000,the United States census by age, gender, race, home ownership, location, city type, statistical area and the recovery rate of census forms were stratified in seven flags, 2010 census still used this method. Switzerland, population census survey of the post-stratification is gender, age, marital status, urban and rural, citizenship, etc

It is obvious that the standards for the signs of these two countries can significantly reflect the missing probability of the individuals in the census and post surveys. If a sign can significantly differentiate the missing probability, it is appropriate to select the sign as a basis for post stratification.

Based on the experience of 2000 census, with the deepening of reform and rapid development of market economy, the change of every aspect for economic and social will continue to deepen and accelerate. The main factors lead to the census omission rate is people's living environment, personnel mobile degree, people's privacy awareness and omissions caused by the interest-oriented issues Due to the temporary and dispersion of the floating population, it brings great difficulty for survey confirmed, which is characterized by a large number of young people out flow and the separation of the city person both lead to higher omission rate. City households separation phenomenon is closely related to the age and marital status.Meanwhile, among the floating population, the proportion of males is higher than females. Based on the regression analysis of data in the omission of 2010 population census of China, the mainly factors are under 20 years of age, especially in aged 0 to 9 children, and then, because of the differences of language, customs and cultural, there are exist many problem during the minority population survey. Thus, according to China's population characteristics, we should choose ethnic, home ownership, degree of urbanization (or rural), population density (densely populated areas, non-densely populated areas), age, gender and so on. These signs cross over-stratification, taking into account sample size, it is necessary to merger some small samples

### Third To further refine the target of post enumeration survey

The target of post enumeration survey should not only estimate total missing rate, but also estimate the error count; not only calculate the population(national) net covering errors (the margin of error between omissions and

errors of the population count is the net covering error), but also calculate the error coverage of different sub-population. Dural- system evaluation requests that the surrey data are correctly registered in the census (recorded in the appropriate locations, within the scope of the census, only recorded once, etc.) of the population.However, the final census population, including except for the right registered population, but also including repetition, fiction, and address errors were incorrectly registered. In order to meet the requirements of the Dural- system evaluation, we should estimate the error count, to clean the final census of the population. It is no doubt that the accuracy of post enumeration survey has a major influence toDural- system evaluation

**Fourth With the Dural- system evaluation estimates the true population of the country**

On the basis of doing well of the three points mentioned above, now we can go straight to the Census's core objective - the use of Dural- system evaluation estimate the national true population. Using the two sets of registration data and population of the post enumeration survey and the popupation in census make up Dural-system evaluation**,**the estimated value:

$$p = \frac{N_c N_p}{N_{11}} \qquad \textit{(NO.1 Formula)}$$

Among it,

P=the true population of the country

$N_c$=the population estimate according to the post enumeration survey

$N_p$=the correct population in census

$N_{11}$=the population being registered twice

But the above formula (Huguihua, 2010) is a theoretical exposition, it cannot calculate in this way in the actual work，because the three variables in the formula cannot be calculated in one step, firstly, you must calculate the population of each stratification, then the sum of all stratification of the population is the true population by country, so that it consistent with the probability of Dural- system evaluation

If record the vth post- stratification of Dural- system evaluation as $Dse_v$：

$$DSE_v = (C_v - D_v) * \frac{N_{cv}}{N_{11v}} * \frac{N_{pv}}{N_{ev}} \qquad \textit{(NO.2 Formula)}$$

Among it,

$C_v$ = the final population of the vth post- stratification

$D_v$ = the personal characteristics delete population of the v th post-stratification

$N_{cv}$ = the correct number of registered population (No. v stratification)

$N_{11v}$ =the population who being registered in the two surveys (No. stratification)

$N_{pv}$= the v th post-stratification population who being registered according to post enumeration survey

$N_{ev}$ = the v th post-stratification population who being registered according to the population census

The meaning of Cv-Dv is exclude the person number that is wrongly registered, and is equal to the population of the post enumeration survey**.** In the last formula (Hogan,H.1993). You will find that the difference bewteenthem is No.2 has one more coefficient

$$\frac{N_{pv}}{N_{ev}}$$

The writer once mentioned that the post sampling suvey data should accurate to the province. So the formula can be independently applied in every province.We can get the true estimated population of one province after collecting all the true population of post-strafication estimated; then,the nationwide population is the sum of all the true population of the provinces.

Two population data can be got by Dural- system evaluation: one is census population, the other is estimated true population. So the missing rate is: net missing rate of one province = (estimated true population – census population of the province)/ estimated true population). Net missing rate of one country = (estimated true popalation – censuspopulation of the country) / estimated true population

## Fifth Keep the independence of population census and post enumeration survey

The independence of population census and post enumeration survey means that the person who participates in the census has no relation with him to participate the post enumeration survey.it also another theory premise for Dural-system evaluation; If the independence do no exist, the Dural- system evaluation would be deviated.

How to keep the independence of population census and post enumeration survey? First, the post enumeration survey should be processed half month later after the end of population census. Because soemthing happened in the population census may affect their attitude towards the post enumeration survey if process at once. Of course, it can't wait too long, otherwise, people would forget their situation on the census day, which may cause the data lost and increase the difficulty of comparability of personal record. Second, try best to employ the othwe persons instead of the census enumerators.we can enroll or train post enumeration survey inspector, or dispatch the excellent census enumerator to the district which different from the population census district.Third, the use of different census methods of investigation. There are three methodsfor population census:the first is to employ the enumerators,the second is by the way of questionnaires, the third is to combine the above two, while Chinese census adopts the first method. If the post enumeration survey use different method, it means all or partial citizen would fill the questionnaire directly and need not face the enumerators. Some countries test to adopt the method door-to-door interviews/door-to-door household survey ,and the post enumeration survey is by mail, online reporting, etc.Sch terms are not mature in China at present, it is still an unsolved problem to collect the data.

## REFERENCES:

1.Wu Jie. The post enumeration survey in census ---The method and research of 2000 census[M]. China Statistics Press. 2000

2.Du Xiaopeng, Hu Guihua. Some suggestions for the post enumeration survey in 2010 China census [J] Northwest Population, 2008 (1)

3.Hu Guahua.The introduction of the post enumeration survey in 2010 America census [J], China Statistics 2009, (8)

4.Hogan, H. The 1990 post- enumeration survey: operation and results[J]. Journal of the American statistical association,1993,88,423,1047- 1060

## RESUME:

Zhou Weizhong,born in September 1957,is the chief captain of Guangxi survey office of the national bureau of statistics, vice-chairman and senior statistician of the eighth board of the national statistical society. He engages in statistical practice, professional statistical theory and economic research of population, agriculture and science, once managed or published many research projects and was awarded outstanding award by the national statistics research progress of science and technology and Guangxi social science research (The third prize).