

Improving the capabilities of survey software for data collection.

Junwoo Jeon

Korea National Statistical Office

920, Dunsan-dong, Seo-gu

Daejeon, KOREA

junwooj@nso.go.kr

1. Introduction

Is it possible to reduce our burden if data collection robots worked with us? Today it is common to send questionnaires to respondents via e-mail, to interview on a doorstep with a notebook computer, and to fill out questionnaires through Internet web pages by respondents. In national statistical survey, these information technologies have already had an their important impact on the method of data collection from the beginning of surveys to published reports.

In this article, I will discuss the change on method of data collection in national statistical survey, improvement of ability in application program, unknown problem in implementation of it and I will suggest an improved model for data collection in the future.

2. Improving in capabilities on data collection method

KNSO (Korea National Statistical Office) uses the PAPI and CAPI, CADI or CSAQ products to collect survey data. It is different from survey to survey, but mainly they have used the mixed method of PAPI and CADI. That is, a field interviewer visits the respondents at their homes or offices and he/she fills out the questionnaire in front of them. They input the collected data via the CADI application program in their offices. The stored data is sent to headquarters by WAN or Internet, and it is analyzed and published.

In 1987, KNSO developed the CADI program for the first time to input handwritten answers. Before, they used the key-entry system to capture questionnaires by keypuncher. This was a vast improvement over previous survey systems and allowed KNSO to process the data faster and introduce quality assurance steps to be sure they had captured the data accurately.

With the increase of personal computers at the beginning of 90' s, the way of data collection has changed from key-entry system to CADI method. Adoption of this new system has made both the developers and researchers suffer from many difficulties such as errors in program development, mistakes in the use of the program and so on. The purpose of this survey application program was to reduce the time of data captures, to minimize coding errors on questionnaire, and to cut total input costs.

In 1994, there was an adventurous attempt on data collection and data processing. It was to change the way of data processing from an IBM Mainframe to a 486 PC on a large scale. The development of the PC is the technical condition for the use of computer technology in large statistical processing. The powerful PC workstation had both processing speed and storage capacity far above that of the mainframes first introduced in production of official statistics.

In spite of a risky attempt, it conducted to process about 2.5 million establishments' data. In the end, we carried out the action of each stage on the whole survey with the application program such as editing, range checking, tabulating, and statistical analysis. Time between survey data collection and publication of the results has decreased greatly.

After that, many functions have been added to the application program for the purpose of improving the data quality and providing the users with convenience. The program was consisted of questionnaire capture, relation checking among items, omission and duplication checking, input range checking, checked error printing, tabulation printing, establishments list managing, and data searching. It's a kind of all-in-one package, which was fully equipped for a survey.

But, the statistical system based on new computer technology has created new challenges to development. This process has provided the program developer with new insights regarding the users of this data. There were many difficulties such as a lack of user's basic computer skills, difficulty of usage on complex menu for a variety of functions, and imperfection of communication systems for renewal application program to the local user.

In order to be successful, the strategy for the statistical system has to overcome the problems above. According to our experience of adopting new survey application systems, especially in the user's computing environment and training, it is important to include computing environment for new system, users education for operation, and convenience of usage to be successful.

In 2001, almost all of survey programs produced by KNSO were contributed to our local interviewer with various functions. It is transformed into the Windows-based program with Delphi, Visual Basic or Power Builder on Windows.

3. Conclusion

We have a tendency to depend upon computers and information technologies in many times. Now, we meet all the requirements of the user's computing environment, developer's developing environment, and communication infrastructure than we did during the initial stages of development. So if we could solve the above problems, we could make a powerful system that could control the statistical survey effectively and help to collect the statistical data improved in quality.

RESUME

We look around the progressive changes in data collection method and the process of the application program for improving quality of statistical data over the past 10 years. And I explain the several points of execution that occurred in the process.