

9 Republic of Korea

Electronic Microdata of the Censuses of the Republic of Korea at the East-West Center, University of Hawaii

Minja Kim Choe

Background

The first census of the Republic of Korea took place in 1949.¹ Only the total population count is available from this census, due to the loss of census forms during the Korean War that broke out in 1950. The second population census took place in 1955 from which a number of tables were produced and published. No electronic microdata are available for these censuses.

Computer processing of data began with 1960 census, and as a result, various forms of microdata in electronic forms exist for this and later censuses. Until 1995, Korean censuses have been collected through interviews by enumerators who visited each household. The completed census schedules were transferred to electronic format by keying in the information for censuses of 1960 through 1985. For the 1990 and 1995 censuses, information in completed schedules was transferred to optical

Dr. Minja Kim Choe is a Senior Fellow in the Program on Population at the East-West Center and adjunct associate professor in Public Health Sciences at the University of Hawaii. Her research interests are fertility and reproductive health; child survival, family and gender issues; and East Asia.

¹ During the Japanese occupation of Korea, censuses were taken in 1925, 1930, 1935, 1940, and 1944.

mark reader (OMR) forms by the enumerators and the OMR forms were processed electronically.

The first public use sample of census data was made from the 1995 population and housing census.² For the censuses of 1960 through 1985, no public use samples were made available. There are, however, sample census data created and made available to selected scholars in Korea and abroad for designated projects.

Microdata Holdings at the East-West Center

The East-West Center is one of the institutions with whom the National Statistical Office has been collaborating on the in-depth analysis of census data for many years.³ The largest project has been the estimation of fertility and mortality from census data. This project includes the analysis of every census since 1960. Population growth has been a major concern of the Korean government from the mid-1960s and throughout 1970s and 1980s. Population planning has been an integral part of the national economic planning since mid-1960. Generating accurate population data in a timely manner, naturally, became an important goal of the national statistical office. With the vital registration system suffering from delayed registration and under-registration (Kwon 1977; Coale, Cho and Goldman 1980), census data provided an excellent alternative source of data for fertility and mortality estimation. A number of methods of estimating fertility and mortality from limited data developed during the 1960s and 1970s, combined with interests of international funding agencies to support basic demographic research, facilitated these projects. Fertility estimation methods developed by Cho and colleagues at the East-West Center (Cho and Retherford 1978; Cho, Retherford and Choe 1986; Luther and Cho 1988; Retherford, Cho, and Kim 1984) and the mortality

² For more information on the these data, see the Korean National Statistical Office web-site at <http://www.nso.go.kr/shopmall/>. The web-site <http://www.nso.go.kr/eindex.html> has some information in English.

³ The National Statistical Office was called the Economic Planning Board from 1960 through 1970, and the National Bureau of Statistics from 1975 through 1985.

estimation methods developed by Brass (1975) and Feeney (1980) were used most frequently. In fact, data from Korean censuses were often used to illustrate some of the newly developed methods (Cho and Retherford 1978; Luther and Cho 1988; Retherford, Cho, and Kim 1984; Retherford and Ogawa 1978). On a smaller scale, analyses of household structure, education, employment, and marriage were carried out (e.g., Choe and Li 1997; Lee and Palloni 1992; Mason 1986; Mason et al. 1986; Zeng et al. 1994). Most of these projects used data from decennial censuses. For these collaborative works, sample files of census data were made available to the East-West Center.⁴

This essay describes electronic microdata files of the censuses of the Republic of Korea that are available at the East-West Center for the years 1960 through 1990 (as of May 15, 2000). The 1995 census data are not included in this essay because the public use sample and its description are readily available from the National Statistical Office of the Republic of Korea. For some censuses, there are more than one sample files of microdata at the East-West Center. For these years, the files with most variables are selected for description in this document.

The East-West Center received sample census data for the years 1960 through 1990 on magnetic tapes. Recently, files on these tapes have been copied onto optical disks for better preservation and faster access. They are now being copied onto CD-ROMs as a part of the IPUMS-International project.

Contents of Microdata Files at the East-West Center

Table 9-1 summarizes the basic information about the most comprehensive sample file for each census currently held at the East-West Center. Table 9-2 shows the list of population variables and Table 9-3 lists housing variables in these files, each by category. With the exception of 1975 census for which we only have a selected number of population variables from the original census forms, the list of variables reflects the variables

⁴ Work by Lee and Palloni took place at the University of Wisconsin-Madison. When Lee joined the East-West Center in 1992, she brought with her the sample census files used for the project.

Table 9–1. Sample Microdata Holdings at the East-West Center of the Censuses of the Republic of Korea, 1960 through 1990^a

Census Year	Sample at East West Center^b	Population Variables	Housing Variables
1960	10 percent	All	All
1966	1 percent	All variables in the long form	Not collected
1970	1 percent	All variables in the long form	None
1975	5 percent	Some variables in the long form	All variables in the long form
1980	2 percent	All variables in the long form	All variables in the long form
1985	2 percent	All	All
1990	2 percent	All variables in the long form	All variables in the long form

^a As of May 15, 2000. ^b Use of all data sets requires authorization from the National Statistical Office in Korea.

collected in each census. For the censuses of 1960 and 1985, no sampling was used and only one enumeration form was used for each census. For other censuses, short forms for most households and long forms were used for samples of households. The files at the East-West Center come from the long forms for these years. Housing variables are available for 1960, 1975, 1980, 1985, and 1990. The 1966 census did not collect housing information. For the censuses of 1970 housing variables are not included in the sample file sent to the East-West Center.

Table 9-2. Population Variables in Korean Microdata Samples at East-West Center

Census Year	1960	1966	1970	1975	1980	1985	1990
Geographical area of residence	X	X	X	X	X	X	X
Age, date of birth	X	X	X	X	X	X	X
Sex	X	X	X	X	X	X	X
Relationship to head	X	X	X	X	X	X	X
Nationality	X						
Education	X	X	X		X	X	X
Marital status	X	X	X	X	X	X	X
Age at first marriage				X	X		X
Migration	X		X		X	X	X
Fertility	X	X	X	X	X	X	X
Mortality		X	X	X	X	X	X
Religion						X	
Handicapped status					X		
Economic activity	X	X	X		X	X	X
Commuting					X		X
Household economics	X	X					
Household composition	X	X	X		X		

Basic variables that are included in all census sample files are: area code of the residence, age, sex, relationship to the head of household, marital status, and fertility. Nationality was collected only in the 1960 census. Education information was collected in all censuses but the file at the East-West Center for 1975 does not have them. Age at marriage was collected only in the 1975 and 1990 censuses. Migration information is available in the censuses of 1960, 1970, 1980, and 1990. Mortality information is available in all censuses except for the 1960 census. Information on handicap status was collected only in the 1980 census. Key variables on economic activity were collected in all censuses except for 1975 and are included in the sample files at

Table 9-3. Housing Variables Included in Sample Census Data Files at East-West Center

Census Year	1960	1975	1980	1985	1990
Type of living quarter	X	X	X	X	X
Ownership/rental status	X	X	X	X	X
Number of households sharing the unit		X	X	X	
Amount of rent/lease				X	X
Length of residence	X				
Total number of rooms in the unit	X	X	X	X	X
Number of rooms used	X		X	X	X
Total floor area			X	X	X
Electricity	X	X			X
Source of drinking water	X		X		X
Type of cooking fuel	X	X	X	X	X
Type of heating	X		X	X	X
Piped water		X	X		
Toilet facility	X		X	X	X
Cooking facility	X		X	X	X
Bathing facility	X		X	X	X
Sharing of facility					X
Household items owned	X	X	X	X	X
Type of building	X	X	X	X	X
Multi-purpose housing unit	X			X	X
Material of wall	X	X	X		X
Material of roof	X	X	X		X
Number of units in the building	X				
Size of land			X		X
Age of the building	X	X	X		X
Condition of the building	X				
Condition of the chimneys	X				
Total maximum number of households for unit					X

the East-West Center. The Korean census began to collect information on commuting in 1980.

Methods of collecting data and coding of the variables vary from census to census for all variables except sex. Geographical codes changed to reflect growth of urban areas, changes in the designation of urban areas, and depopulation of some rural areas. For example, in 1960 there were 10 large administrative units in Korea consisting of one metropolitan city and 9 provinces. Thirty years later, there were 6 metropolitan cities and 9 provinces. At the East-West Center, documentation is available for some of the changes. Full documentation is in the process of compilation.

In Korea, age is a very important part of personal identification together with name and clan. Most people have accurate knowledge of their own age as well as the ages of their family members, close relatives, and close friends. However, ascertaining age in a uniform manner is not a simple matter because two different calendar systems—Chinese lunar and Western solar—have been in use for a long time in Korea. Some people remember important dates according to the lunar calendar and others according to the solar calendar. The complication associated with two calendar systems was not given due attention in the 1960 census. Not enough information was collected to ascertain if the reported age was by lunar calendar or the solar calendar (see Coale, Cho and Goldman, for example). In later censuses, a carefully designed set of questions was used for the date of birth and age. Therefore, ascertaining age in completed years (age at last birthday) accurately for most people became possible for 1966 and later censuses. A similar level of attention was not given, however, for ascertaining the age at marriage when this variable was included in the census schedule.

The coding of relationship to head of household differs slightly from census to census. One of the main concerns associated with this variable has been how accurately mothers of young children can be identified, and how members of different nuclear families within a household can be identified. The most frequently used method of estimating current and recent fertility in Korea has been the “own children” method (Cho, Retherford and Choe 1986). The accuracy of the estimation depends, in part,

on how accurately the matching of young children and their mothers is made. In Korea, the terms used for relationships between family members and close relatives describe the nature of how two individuals are related through combinations of sibling relationship, marriage, and detailed parental lineage. This precise information on relationships was routinely collected in the censuses but the coding of relationship, however, varied somewhat from census to census.

For most censuses, education was limited to two variables: highest level of schooling and attendance. If the person was no longer attending school, the person was asked whether the level of schooling was completed or not. For earlier censuses (1960, 1966 and 1970) a question on literacy was also asked.

Two additional variables were collected in the earlier censuses but not in later censuses. One is nationality, which was collected only in the 1960 census. For later censuses, non-citizens were enumerated using specially designed forms and were tabulated separately. "Economic activity" of the household (farming or non-farming) was collected in the 1960 census and generated from economic activities of individual members of the household for the 1966 census. Later censuses do not have this variable.

Regarding migration, place of birth is available for 1960, 1970, 1980, 1985, and 1990. Residence one-year before the census is available for 1980, 1985, and 1990. Residence five-years before is available for every census starting with the 1970 enumeration.

On fertility, the total number of children a woman has ever borne is available in all censuses, and by sex for the years 1975 through 1985. For 1966, the number of children born during the 12-month period preceding the census is also available. For the years 1970, 1975, and 1980, number of children living with their mothers is also available. This information is used for the process of matching women with their own children so that the estimation of age-specific fertility would be more accurate.

Mortality information is available in the form of a combination of two variables: total number of children a woman has ever borne and among them, the numbers of children who have died. This information is available for all years except 1960 and also by sex of children for the years 1970 through 1985. For

1966, the information is also available for children born during the 12-month period before census. See the Appendix 9-1 for details on how the variables on fertility and mortality were collected in each census.

Three basic variables on economic activities—the main activity during the reference period, industry of employment, and occupation—are available for all censuses except for 1975. The reference period used for censuses up to 1980 is the week preceding the census date. For 1985 and 1990, “usual” economic activity was asked without specifying the reference period. Class of work is available for the years 1960, 1966, 1970, and 1980. Whether the individual earned any income from work is available for 1960, 1970, and 1980. Additional questions regarding unemployment and underemployment were collected in 1960 and 1970. In the 1960 census, if a person’s main activity was “unemployed” and was not looking for work, he/she was asked why not. In the 1970 census, number of months a person was working during the year before the census was asked. More details on these variables are available in the Appendix 9-2. Coding of industry and occupation changed many times during the 1960-90 period reflecting rapid industrialization of the economy. Documentation on these changes is available at the East-West Center.⁵

Housing variables have changed substantially over the years, as shown in Table 9-3. In Korea, renting a part of a housing unit has been common and the housing censuses have collected information on the sharing of housing units and the availability of basic facilities such as kitchen, toilet, and private entrance. Questions used vary slightly from census to census. As residential buildings with multiple units became more common in recent years, housing censuses collected additional variables on characteristics of buildings as well as housing units. An interesting section of the housing census is the list of household appliances (Appendix 9-3). Changes in the list reflect changes in living conditions, energy use, life style, access to mass media, and transportation in Korea during the 1960-90 period.

⁵ They are also available at web-sites <http://www.nso.go.kr/shopmall/> and <http://www.nso.go.kr/eindex.html>.

Completeness of Enumeration

The enumeration of the Korean population has been interview-based until the 1995 census and the completeness of enumeration has been estimated to be very good. Most census enumerations were followed by post-enumeration surveys (PES). The overall levels of underenumeration estimated from these are small, in the range of 1.0 – 2.0 percents for most censuses and never exceeding 5.0 percent (National Bureau of Statistics 1982b; National Statistical Office 1991).

Confidentiality

The sample census files do not have individual names. The residence of individuals can be identified only up to the third level of administrative division of the country (Dong, Eup, or Myon). For the 1990 census, the number of the third-level administrative units was 3,505. The average population size for an administrative unit was 12,385. Thus, the confidentiality of individuals is ensured for most applications. The author is not aware of any other specific actions that have been taken to assure the confidentiality of individuals.

Data Access

As mentioned above, the microdata files at the East-West Center were created for specially designated research projects, with the understanding that the files would be used only for those projects. The East-West Center does not have the authority to make the sample census data available to potential users. In general, approval of the National Statistical Office is required for any use of these files by individuals or institutions who are not affiliated with the previously agreed-upon projects.

Documents at East-West Center

For each sample census file at the East-West Center, we have complete documentation including record layout and coding instructions for each variable in the file. As mentioned earlier, coding of some variables changed from census to census,

and we have complete documentation of the changes except for changes in geographical codes between 1970 and 1990. We are in communication with the National Statistical Office to obtain a set of full documentation.

Research Possibilities

The set of census sample files spans more than 30 years. This period was characterized by rapid economic development, social change, and demographic transition and the Korean census microdata provide excellent research opportunities on a variety of topics. The most obvious research topic is the in-depth analysis of population dynamics—including the analysis of fertility and mortality—and their consequences on population age structure, dependency ratios, population aging, and labor force dynamics. Population movements can be analyzed together with socioeconomic changes of the sending and receiving communities.

Korea has experienced remarkable economic development during the 1960-90 period. The census data provide valuable resources for the examination of economic and social changes for the whole country as well as regions. Statistics on variables such as education, occupation, and industry from population censuses, and variables such as ownership of housing, size of living areas, basic housing facilities, and ownership of major household items from housing censuses may be used to develop indicators of socioeconomic conditions, exposure to mass media, and life styles.

Characteristics of communities in terms of the individuals and households within them can be generated and analyzed. The range of community characteristics that can be generated from census data is wide. Distribution of occupations and educational levels can be used to reveal much about socioeconomic conditions of communities. Sex ratio of children under age 5 and the proportion of women with young children who are employed in the non-agricultural sector can be generated as indicators of status of women in the community. Variations in proportion of the elderly who live with their children indicate one impact of migration. Possible research projects using community data include the study of

interrelationships between economic development and social change, causes and consequences of the growth of urban areas, and population concentration at select metropolitan cities.

Some measures of environmental conditions of communities can be generated from census data and analyzed, together with demographic and economic data. Indicators of environmental conditions can be generated from housing data on fuel, toilet facility, car ownership, and means of commuting, among others. Thus, the environmental impact of population dynamics and economic development can be studied to a certain extent.

Publications

A number of publications resulted from research based on census data, carried out at the East-West Center or elsewhere by staff members of the Center. Most extensive are those estimating fertility and mortality, including the development of new measures and methods (Cho 1971; Choe and Retherford 1978; Coald, Cho, and Goldman 1980; Luther and Cho 1988; Retherford, Cho, and Kim 1984; Retherford and Ogawa 1978). Others include analyses of household structure and their projections, and marriage behavior (Choe and Li 1997; Lee and Palloni 1992; Mason 1986, Zeng et al. 1994).

Expert Users

Lee-Jay Cho led the fertility analysis with Robert D. Retherford, Griffith Feeney, Minja Kim Choe, and Norman Y. Luther at the East-West Center as well as population experts outside the Center including Nam-Il Kim (Dong-A University), Ansley Coale (Princeton University), and Noreen Goldman (Princeton University). For the analysis of household structure, education, employment, and marriage, Yean-Ju Lee, Andrew Mason, and Minja Kim Choe at East-West Center collaborated with colleagues elsewhere (Aberto Palloni and University of Wisconsin, Sung Yeal Koo at Yonsei University, Zeng-Yi at Beijing University, and Li Lei at University of Hawaii).

Bibliography

- Brass, William. 1975. *Methods for Estimating Fertility and Mortality from Limited and Defective Data*. Chapel Hill: International Program of Laboratories for Population Statistics, University of North Carolina.
- Cho, Lee-Jay. 1971. "Korea: Estimating current fertility from the 1966 census," *Studies in Family Planning* 2: 74-75.
- Cho, Lee-Jay and Robert D. Retherford. 1978. "Own children fertility estimates by duration since first marriage: Preliminary results for Cheju Province, Republic of Korea," *Asian and Pacific Census Forum* 5(1): 6-10.
- Cho, Lee-Jay, Robert D. Retherford, and Minja Kim Choe. 1986. *The Own-Children Method of Fertility Estimation*. Honolulu: East-West Center.
- Choe, Minja Kim and Lei Li. 1997. "Later Marriage or No Marriage?: An Examination of Marriage Patterns in Korea Using a Mixture Model for Duration Data," paper presented at the 23rd General Population Conference of the International Union for the Scientific Study of Population, 11-17 October 1997.
- Coale, Ansley J., Lee-Jay Cho, and Noreen Goldman. 1980. *Estimation of Recent Trends in Fertility and Mortality in the Republic of Korea*. Washington, D.C.: National Academy of Sciences.
- Economic Planning Board. 1963. *1960 Population and Housing Census of Korea*. Seoul.
- _____. 1969. *1966 Population Census Report of Korea*. Seoul.
- _____. 1973. *1970 Population and Housing Census Report*. Seoul.
- Kwon, Tai-Hwan. 1977. *Demography of Korea: Population Change and Its Components 1925-66*. Seoul: Seoul National University Press.
- Kwon, Tai-Hwan, Tai-Hun Kimn and Chin-Ho Choe. 1995. *Population and Family in the Republic of Korea (in Korean)*. Seoul: Ilsinsa.
- Lee, Yean-Ju and Alberto Palloni. 1992. "Changes in the family status of elderly women in Korea," *Demography* 29(1): 69-92.
- Luther, Norman Y., and Lee-Jay Cho. 1988. "Reconstruction of birth histories from census and household survey data," *Population Studies* 42(3): 451-72.
- Mason, Andrew. 1986. *HOMES: A Household Model for Economic and Social Studies*. Papers of the East-West Population Institute. Number 106. Honolulu: East-West Center.
- Mason, Andrew, Daniel B. Suits, Sung-Yeal Koo, Naohiro Ogawa, Mathana Phananiramai, and Hananto Sigit. 1986. *Population growth and economic development: lessons from selected Asian countries. Policy Development Studies, No. 10*. New York: United Nations Fund for Population Activities.

- National Bureau of Statistics. 1977. *1975 Population and Housing Census Report*. Seoul.
- _____. 1982a. *1980 Population and Housing Census Report*. Seoul.
- _____. 1982b. *Evaluation of Recent Korean Censuses*. Seoul.
- _____. 1987. *1985 Population and Housing Census Report*. Seoul.
- National Statistical Office. 1991. *Population Projections: 1990-2021*. Seoul.
- _____. 1992. *1990 Population and Housing Census Report*. Seoul.
- _____. 1996. *Population Projections*. Seoul.
- _____. 1997. *1995 Population and Housing Census Report*. Seoul.
- Republic of Korea, National Bureau of Statistics. 1984. *The Levels and Trends of Fertility for Small Geographical Areas in Korea*. Seoul.
- Retherford, Robert D., Lee-Jay Cho, and Nam-Il Kim. 1984. "Census-derived estimates of fertility by duration since first marriage in the Republic of Korea," *Demography* 21(4): 537-74.
- Retherford, Robert D., and Naohiro Ogawa. 1978. "Decomposition of the change in the total fertility rate in the Republic of Korea, 1966-1970," *Social Biology* 25(2): 115-27.
- Zeng Yi, Ansley Coale, Minja Kim Choe, Liang, Zhiwu, and Liu, Li. 1994. "Leaving parental home: census-based estimates for China, Japan, South Korea, United States, France, and Sweden," *Population Studies* 48(1): 65-80.

**Appendix 9-1. Fertility and Mortality Information
Collected in the Population Censuses
of the Republic of Korea, 1960-1990**

	1960	1966	1970	1975	1980	1985	1990
Number of children ever borne (CEB)	X	X	X	X	X	X	X
CEB by sex			X	X	X	X	
CEB during last year		X					
Number of children living (CL)		X	X	X	X	X	X
CL by sex			X	X	X	X	
CL during last year		X					
Number of children who died (CD)		X	X	X	X	X	X
CD by sex			X	X	X	X	
CD among those born during last year		X					
CL living with mother			X	X	X		
CL living with mother by sex			X	X	X		

Note: The information was collected from all ever-married women in the 1960 census and ever-married women age 15 or older in later censuses.

**Appendix 9–2. Information on Economic Activity
Collected in the Population Censuses of
the Republic of Korea, 1960-1990**

	1960	1966	1970	1980	1985	1990
Reference period	1 week	1 week	1 week	1 week	usually	usually
Main activity	X	X	X	X	X	X
Industry	X	X	X	X	X	X
Occupation	X	X	X	X	X	X
Class of work	X	X	X	X		
Worked for money	X		X	X		
Was looking for work?	X	X		X	X	X
If not, why not?	X					
Months worked last year			X			

Note: The information was collected from men and women age 10 or older in 1960 census and age 14 and older in later censuses.

Appendix 9-3. Items Included in the List of Household and Appliances in the Housing Censuses of the Republic of Korea, 1960-1990

	1970	1975	1980	1985	1990
Sewing machine	X				
Radio	X				
TV	X	X	X		X
TV, color				X	
TV, black and white				X	
Audio system	X	X			X
Telephone	X	X	X	X	X
Refrigerator	X	X	X	X	X
Piano or organ	X	X	X	X	
Newspaper subscription		X	X		
Washing machine		X	X	X	X
Air conditioner			X	X	X
V.T.R.				X	X
Car				X	X
Personal computer					X

