



The Accuracy and Coverage of Internet Based Data Collection for Korea Population and Housing Census

2009. 3.

Korea National Statistical Office



1

Introduction

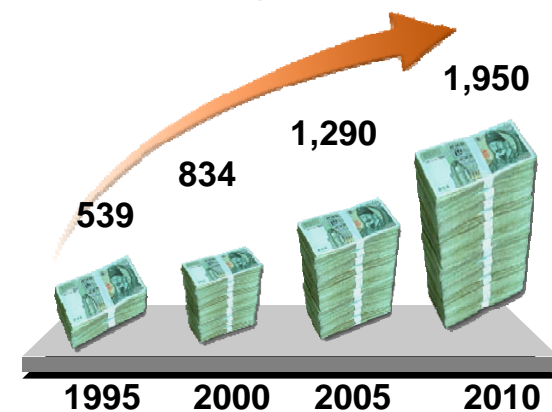
❖ Outline of the 2010 Census

- ❖ 5 years basis since 1925
- ❖ Census-taking period : November 1~26
- ❖ Census Items : 52 (short form 19, long form 33)
- ❖ Enumeration methods : Internet, Mail, Interview
- ❖ Data capture : Intelligent character recognition (ICR)
- ❖ Release of results : 2010.12~2011.12

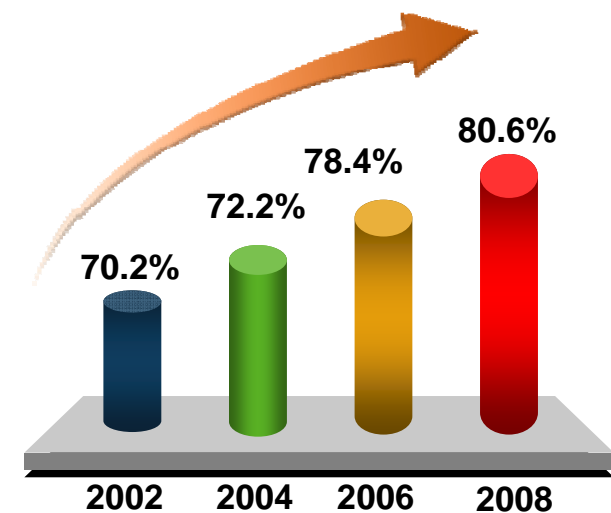
❖ Environment of Census-taking

- ❖ Burden of a huge budget
- ❖ Increase of ageing people and one-person household
- ❖ Increase of the Internet access rate
- ❖ More likely to use administrative data
- ❖ Growing awareness of privacy

Census budget (billion won)



Internet access rate



2 Data Collection Methods

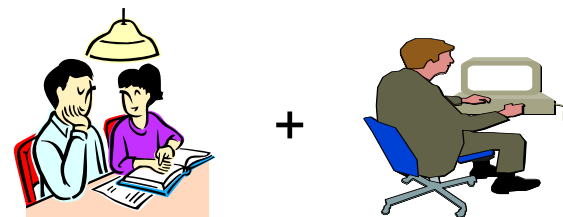
◆ Before 2005

Enumerator method



◆ 2005 Census

Enumerator method
+
Internet (0.9%)



◆ 2010 Census

Internet (30%) Mail (20%) Enumerator (50%)





Pull & Push Strategy for 30% of Internet participation

Pull Strategy

Public campaign

Campaigns for Internet, leaflet for students

Incentives

Gift certificates by lot

Convenient system

Internet access code

Push Strategy

Internet First

Internet → Mail → enumerators

Extend period

Open Internet system until the end of census-taking period

Without questionnaire

A letter asking participation via Internet

3

Accuracy of Internet Method

Mode effect of
Internet Method

Accuracy of data

Coverage among all answers

Indicators for accuracy of data

- ◇ Respondent error
- ◇ Processing error : data capture error, edit failure rates
- ◇ Non response error
- ◇ Coverage error



Advantages of Internet method for data accuracy

- ◆ User-friendly questionnaire form
- ◆ Online checking of completeness
- ◆ Automated skips and user guidance
- ◆ Interactive aids (flash up, warning messages) and explanations



Respondent Error

Table 1. Correspondence rates between the 2005 census and the post-enumeration survey

	Interview	Internet
Age	98.7%	99.0%
Relationship to the head of households	99.3%	99.1%
Marital status	98.9%	99.9%

- ◇ **Internet** : Answer more frankly to questions related privacy (marital status, age)
- ◇ **Interview** : Advantages in complicated questions (relationship to the head of households)



Non-Response Error

Table 2. Item non-response rates by data collection methods

Item non-response rates	Total	Interview	Mail	Internet
Total	1.7 %	2.1%	2.2%	0.01%
Question for household member	2.3%	2.9%	2.9%	0.02%
Question for household	1.0%	1.2%	1.3%	0.00%
Question for house	0.1%	0.1%	0.1%	0.00%

- ◇ Internet : item non-response rates close to 0.0%
 - Check box, Number : close to 0.0%
 - Alphabet : Job (3.2%), Occupation (1.1%)



Coverage Error

Table 3. Coverage error for household members

	All respondent	Respondent via internet
Missing rates	1.49%	0.4%
Duplication rates	2.39%	0.8%
Net coverage rates	0.90%	0.4%
Total Coverage rates	3.88%	1.20%

* Foreigners' coverage rates are not included

* Data source : Internet method evaluation survey for the 2005 Census

- ◇ Internet : fewer missing and duplicated answer
 - User friendly explanations using illustrations
 - Characteristics of participants : the young, high educated person



Processing Error

◇ Edit failure rates

Table 4. Edit failure rates per household of the 1~3 pre-test
(unit : case)

Total	Internet	Interview	Mail
2.0	1.2	1.9	2.4

◇ Data capture error rates of the 2005 Census : 0.19%



4

Coverage of Korean Census on the Internet

Table 5. Coverage of Internet among all census answers
(Internet penetration rates)

(Unit : %)

2005 Census	1 st pre-test	2 nd pre-test	3 rd pre-test
0.9	13.3	3.9	22.1

- ◇ 2nd pre-test : Internet & mail → Enumerator
- ◇ 1st & 3rd pre-test : Internet → Mail → Enumerator



Table 6. Coverage of respondents via Internet among all respondents by age group

(unit : %)

Age	Coverage
Total	15.5
Under 10	20.0
10-19	19.0
20-29	17.2

Age	Coverage
30-39	17.0
40-49	16.6
50-59	12.4
60+	7.3

◇ **Younger persons tends to respond more via Internet**



Advertising campaign in schools



Gaps of Internet accessibility between the young and the old



Table 7. Coverage of respondents via Internet among all respondents by type of household

(Unit : %)

Type of household	Coverage	Type of household	Coverage
Total	17.0	One-person households	5.6
One-family household	16.6	Households with no blood ties group	4.1
Household consisted a family and non-family members	12.4	-	-

- ◇ **Low coverage rates in hard-to-enumerate groups**
 - One-person households (5.6%)
 - Households with no blood ties (4.1%)



◇ Coverage rates by type of house

Apartment	Ordinary house
18.0%	6.1%

◇ Coverage rates by type of occupancy of house

Owned	Rent
14.7%	11.0%

5

Conclusion

Mode effect
(Internet method)

Increase overall data quality

Coverage gaps by
participants' characteristics

Negative effect on time
series analysis



Thank You

