

---

# 2010 인구센서스 데이터 품질관련 지역과정 참가 결과 보고

---

2009. 11.

I. 회의 개요 .....	1
II. 세부 일정표 .....	2
III. 회의 주요 내용 .....	3
IV. 시사점 및 참가소감 .....	14

<참고자료 : 싱가포르 및 일본의 센서스 개괄>

인구조사과



## I. 회의 개요

□ 일시 및 장소 : 2009년 10월 18일~24일 (5박 7일), 중국 마카오

□ 회의명 : 2010 인구센서스 데이터 품질관련 지역과정

**(Sub-regional Course on 2010 Population Census Data  
Quality 19 - 23 October 2009, Macao, China)**

□ 참가국 및 인원

○ 참가자 : 인구조사과 황수린주무관

○ 아시아 태평양 지역 10개국 총 20명 참석

- 한국, 중국, 싱가포르, 홍콩, 말레이시아, 인도네시아, 베트남,  
브루나이 대표단 : 16명, 마카오통계청 : 4명

- 20명 중 14명이 여성 (참가신청시 여성 우대 명시함)

□ 주요 회의내용

○ 센서스 품질평가 소개

○ 품질평가 기준과 데이터 입력방법

○ 내용검토와 결과 공표 등

□ 동 회의의 특징

○ 각국의 2010 센서스 관계자들의 주제별 발표 및 적극적인 질문과  
토론형식

○ 1국가당 1개 이상 Country report 작성 및 발표

○ 역할연극(Role play) : 조사원 vs 응답자, 조사표설계자 vs 조사원 등

○ UNSIAP 및 마카오 통계청에서 숙박 및 식비(조식, 중식)제공

## II. 세부 일정표

구 분		시 간 대	주요내용
1일차	10. 19. (월)	09:00~09:30 09:30~12:00 14:00~16:00 16:00~17:30	<ul style="list-style-type: none"> <li>· 등록</li> <li>· 환영사 및 과정 소개 등</li> <li>· 인구센서스의 평가 개괄               <ul style="list-style-type: none"> <li>- 각 단계에서의 오류</li> <li>- 다른 자료와의 비교, 지역적분석, 사후조사등</li> </ul> </li> </ul>
2일차	10. 20. (화)	09:00~09:30 09:30~12:00 14:00~16:00 16:00~17:30	<ul style="list-style-type: none"> <li>· 인구센서스의 평가방법               <ul style="list-style-type: none"> <li>- 각 국의 경험 공유</li> </ul> </li> <li>· 자료수집 방식 소개               <ul style="list-style-type: none"> <li>- 커버리지 에러, 응답에러 측정에러등</li> <li>- 응답부담을 줄이는 방안 검토</li> <li>- 각국의 경험 공유</li> </ul> </li> </ul>
3일차	10. 21. (수)	09:00~09:30 09:30~12:00 14:00~16:00 16:00~17:30	<ul style="list-style-type: none"> <li>· 자료수집 방식 소개 (계속)               <ul style="list-style-type: none"> <li>- 새로운 자료수집 방식</li> <li>- IT 기술을 이용한 자료수집의 장/단점</li> </ul> </li> <li>· 자료 내검 (editng)               <ul style="list-style-type: none"> <li>- 무응답에 대한 처리방법, 가중치 조정등에 관한 각국의 기술 공유</li> </ul> </li> </ul>
4일차	10. 22. (목)	09:00~09:30 09:30~12:00 14:00~16:00 16:00~17:30	<ul style="list-style-type: none"> <li>· 센서스 결과 공표               <ul style="list-style-type: none"> <li>- 각국의 결과공표 정책</li> </ul> </li> <li>· 전통적인 결과공표방식에서 ICT기술을 이용한 방식으로 변화               <ul style="list-style-type: none"> <li>- 각국의 경험과 문제점</li> </ul> </li> </ul>
5일차	10. 23. (금)	09:00~09:30 09:30~12:00 14:00~16:00	<ul style="list-style-type: none"> <li>· 소지역통계와 GIS를 이용한 기술               <ul style="list-style-type: none"> <li>- 일본의 경험</li> </ul> </li> <li>· 과정 평가시간</li> </ul>

### III. 주요 회의 내용

#### 1. Course 소개

(Mr. H. Furuta : Course Director)

##### □ 참가 자격 및 목표

- 2010 센서스 업무에 관련된 중간이상 경력의 통계공무원 (여성 우대)
- 각국의 경험공유와 강의를 통해 2010 센서스에서 효율적인 자료 수집과 데이터 프로세싱, 적절한 자료제공이 가능하게함

##### □ Course 스케줄 및 회의장소

- 총 18개 세션, 하루 4개 세션 (각 세션당 75분 소요)
- 마카오 Grand Lapa Hotel, Lotus Room

#### 2. 2010 센서스 데이터 품질 개괄

(Mr. Nath UNSIAP Consultant)

##### □ 센서스 품질이란 무엇인가

- National Quality Assurance Framework에 따르면 품질이라는 용어를 넓은 의미로 다음과 같이 설명하고 있음
- “통계적 절차들이 얼마나 잘 수행되고 있는지, 결과가 이용자와 업무 담당자의 기대를 만족시켜주는지를 포함한다. 좋은 품질이란 이용자의 수요를 만족 시켜주는 것 뿐만 아니라 응답부담과 개인정보보호, 목표, 비용효과적인 절차가 만족되는 것도 중요함”

## □ 품질에 대한 국제적 노력

- 1994년 UN Statistical Commission (UNSC)은 '공식통계기관의 기본원칙'을 선포함 10가지 원칙 중 어떤 것도 명시적으로 데이터 품질을 언급하지는 않았지만 모두 기본적으로 품질관리의 기본이 되는 것임
- 2005년에 UNCCSA는 국제기관이 통계를 축적하기 위한 비슷한 원칙을 언급함 UNSD는 UNSC의 38차 세션에서 데이터품질을 관리하는 것에 대해 패널 토의함

## □ 품질에 관한 6가지 요소

- 관련성(Relevance)
  - 이용자들의 수요를 만족시키는 정도를 반영함
- 정확성(Accuracy)
  - 통계정보가 측정하고자하는 현상을 정확하게 기술하는 정도를 의미
    - \* 좁은 의미의 데이터 품질
    - \* 관련사항 : 커버리지, 샘플링, 무응답, 자료입력, 코딩, 비표본오차등
- 시의성(Timeliness)
  - 기준이 되는 시점과 정보이용가능시점간의 차이가 적절한지를 의미
    - \* 자료제공 정책, 조사주기
- 접근가능성(Accessibility)
  - 통계정보가 존재한다는 사실을 확인하고 정보에의 접근이 쉬운 정도
- 이해가능성(Interpretability)
  - 부가적인 정보와 메타데이터가 이해가능 해야하고 이용할 수 있어야함
    - \* 이는 메타데이터의 한 부분이며 조사표 디자인 단계에서부터 고려되어야 할 사항임
- 일관성(Coherence )
  - 일반적인 통계분석틀 안에서 다른 통계정보들과 접합이 가능한지

□ 다음의 틀이 앞서 언급한 6가지 요건의 지표가 될 수 있음

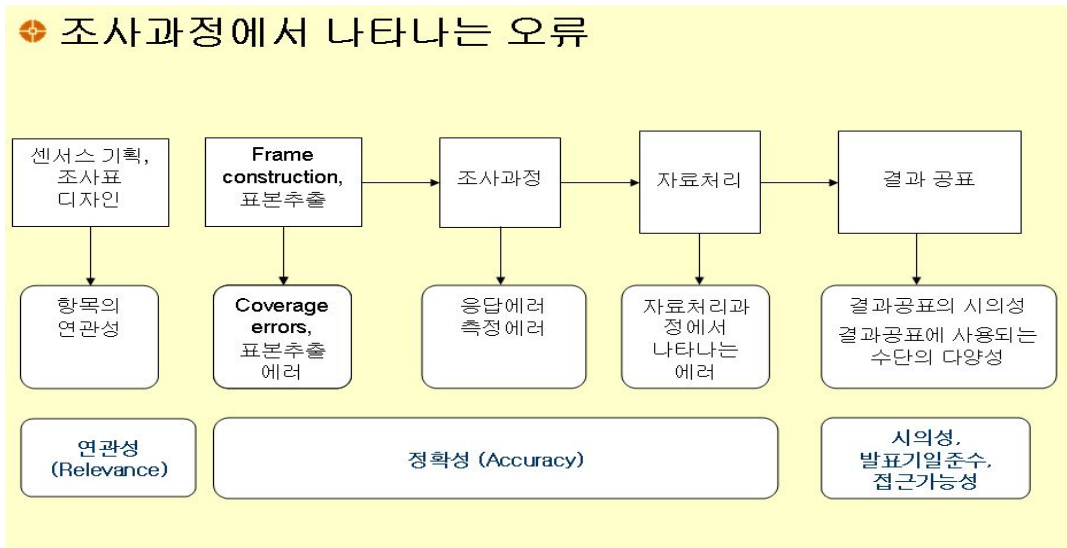
- 지난 센서스의 경험
- 성인지 자료
- 센서스 기본틀을 설정하는 단계에서 검토
- 효율적인 데이터에디팅 시스템
- 다른 자료로부터 유효성 검사
- 사후조사

### 3. 2010 센서스 데이터 품질 구성요소

(Mr. H. Furuta. Course Director)

□ 품질의 구성요소

- 결과물과 관련된 요소 : 관련성, 일관성, 비교가능성
- 과정과 관련된 요소 : 정확성, 시의성, 접근성, 명확성



## □ 오류의 종류

### ○ 표본에러

통계이론을 이용한 통제, 룱폼, 샘플데이터를 이용한 집계

### ○ 비표본에러

- 커버리지에러, 내용에러
- 조사전후 통제가능성
- 가능한 에러를 예측하고 에러의 원인을 파악
- 이전센서스에서 경험
- 다른 국가에서의 경험

## □ 조사기간 이후 오류를 측정하는 방법

### ○ 사후조사의 효용성

- 센서스 기획자에게는 다음 센서스를 개선시키는데 필요한 정보를 줌
- 센서스 데이터 이용자에게 센서스 결과의 품질에 대한 정보를 제공함

### ○ 다른 정보와 비교

주민등록자료나 사회보험 자료 등 집합되거나 개인적인 자료를 비교

### ○ 인구학적 분석

지난 센서스의 결과와 비교하고 출생사망 이민등 자료로부터 얻은 정보와의 일관성을 비교함

## □ 커버리지 에러 (Coverage error)

- 전체 인구나 주택의 수를 세는데 있어서 누락되거나 잘못 포함시킨 에러를 의미함

## ○ 거처 누락

- 거처의 누락은 자동적으로 가구와 가구원의 누락을 야기하므로 센서스에서는 치명적인 에러임
- 거처 누락의 원인
  - ① 지도의 문제점
  - ② 조사원의 문제점 : 지역의 문제, 인구밀도가 매우 높은 지역 부주의 등으로 인한 문제

## ○ 가구 누락

- 가구의 누락은 더많은 보살핌과 관찰이 필요한 가구가 누락되는 경향이 있음 가구의 누락은 자동적으로 가구원의 누락으로 연결되므로 주의가 필요함
- 가구 누락의 원인
  - ① 한거처 이상의 집에서 머무는데 들다 조사가 되지 않음
  - ② 1인가구
  - ③ 조사기간에 거처를 바꾸는 경우
  - ④ 호텔이나 시설에 있는 사람들
  - ⑤ 다른 누락의 원인으로는 같은 거처에 사는 사람이지만 생계를 달리하여 다른 가구로 분리해야 하는데 지침을 잘못 알고 있어서 이를 누락시킨 경우도 있다.

## ○ 가구원 누락

- 원인 : 조사원의 부주의, 조사대상이 되는 가구원에 대한 명확한 지침 부족 병원이나 짧은 여행을 다녀오는 경우, 가구에 사는 구성원과 혈연이 아닌 고용인 등은 응답자에 의해 누락되기 쉬움 새로 태어난 아기도 누락되기 쉬움



## ○ 중복

- 중복은 누락보다는 훨씬 적어서 중복과 누락을 합친 순누락률로 인해 인구는 과소계상됨
- 중복의 원인
  - ① 조사구 지도가 오버랩핑된 경우
  - ② 누락의 원인과 유사한 이유로 중복이 가능 예를 들면 병원에 있거나 여행을 간 가구원이 두 번 계상될 수 있음
  - ③ 조사기간이 긴 경우 중복확률이 높아짐

## □ 내용에러 (Contents error)

- 내용에러는 가구나 가구원의 특성에 대한 상세 정보에 대한 오류를 의미함
- 커버리지 에러가 전체 인구, 가구등 크기에 영향을 주는 반면 내용에러는 결과의 품질에 영향을 줌
- 원인
  - 조사표의 문항이 불충분한 설명이나 응답자로부터 모호한 정보를 얻은 경우 내용에러가 나타남
  - 답변하기 곤란한 민감한 질문들
  - 불충분한 교육
  - 잘못 디자인된 조사표 등

(응답하기에 불충분한 응답란이나 코딩칸이 분리되어 있는 것 등은 응답자가 잘못된 표기를 하지 않도록 하는 중요한 요소임)

- 항목 중에서는 나이와 경제활동상태에 관한 질문에서 에러가 많이 나타남 나이는 지역마다 세는 방식의 차이로 인한 것일 수 있고 경제활동상태는 조사대상 가구원이 아닌 다른 사람이 응답할 경우 잘못 응답할 수 있기 때문임

#### 4. 조사 수행방법

(Mr. Nath UNSIAP Consultant)

##### □ 2010센서스에서 DATA COLLECTION 전략

- 각 가구를 개별적으로 만나서 조사하는 전통적인 방식  
(de Defacto or Dejure 방식 결정)
  
- 등록정보를 활용하는 방법
  - 인구등록정보나 기타 정보를 이용하여 센서스를 수행함
  - 가구에 대한 정보를 미리 조사표에 기입함
  - 장점 : 응답부담을 줄이고 정보처리 단계를 단축시킬 수 있으며 응답편차를 조정하는 데 도움이 됨
  - 한계 : 규모가 큰 센서스에서는 적용하기 어려움
  
- ICT기술을 이용한 방식(인터넷 조사등)
  - 노트북, 자동응답 인터뷰 방식, 인터넷조사

##### □ DATA COLLECTION 단계에서의 비밀보호와 보안

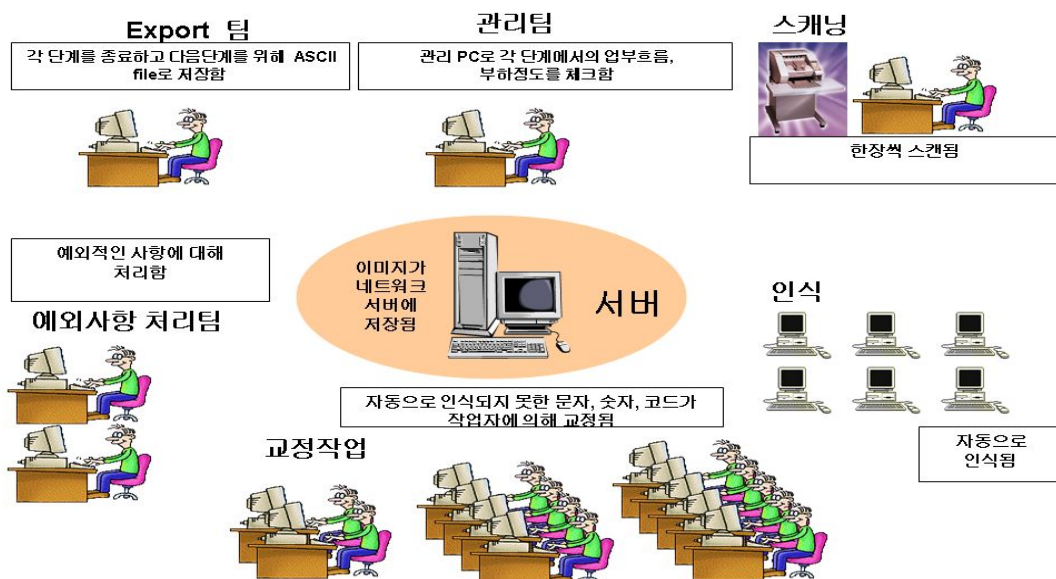
- 정보는 엄격하게 비밀이 보호되며 완벽한 물리적 보안과 개인정보 노출이 없다는 것이 대중에게 확실히 해주어야함
- 단지 승인된 경우만 개인정보에 접근할 수 있음
- 센서스관련자와 서비스를 제공하는 용역업체들은 관련 법규정을 준수해야함
- 센서스 데이터를 다루는 컴퓨터 시스템은 보안이 엄격하고 승인되지 않은 접근은 불가능함

## □ 조사 수행 단계에서의 오류 통제

- 조사원 교육부담, 조사원의 업무과중 등  
통제방법 : 조사원 교육의 질 향상, 강화된 지도감독
- 응답부담 : 너무 많은 항목 수로 인한 응답부담은 핵심항목위주로 항목수를 조정하거나 등록자료를 이용하는 방법이 있음
- 센서스에서 응답한 개인정보가 행정적인 목적으로 사용되지 않을까 하는 심리적 불안감이 있음 이것은 무응답이나 측정에러를 야기함
- 홍보부족은 무응답과 측정에러의 주요원인임
- 법이 기반이 된 강력한 리더십은 센서스의 오류를 줄이는데 기여할 수 있음

## □ 자료입력방법 (DATA capture)

- ICR 방식
  - 대용량 자료처리를 위한 시스템으로 많은 국가에서 지난 센서스에서 적용함 매우 빠르고 결과도 신뢰할만함
  - 다른 조사에 연속적으로 쓰일 수 있다면 매우 비용효과적임



## □ 데이터 내검

- 정확한 센서스 결과를 얻기 위해서는 특히 자료처리단계가 끝난 후에 데이터에 오류와 비일관성이 없어야함
- 자료처리과정에서의 두 가지 오류 유형 「유엔권고안」
  - ① 더 이상의 진행을 막는 오류
  - ② 논리적인 흐름이나 다음 단계의 진행을 막지는 않지만 유효하지 않거나 비일관적인 결과를 낳는 오류
- \* 유엔권고안에서는 첫 번째 오류는 반드시 수정되어야 하고 두 번째 오류는 가능하면 수정되어야함. 따라서 센서스 내검의 기본 목적은 가능한 한 많은 오류를 찾아내어 유효하고 일관성 있는 데이터셋을 만드는데 있음
- 한계점  
자료처리과정에서 모든 유형의 오류를 교정할 수는 없음  
(예: 조사표 내부적으로는 일관되어 보이지만 응답자가 거짓으로 대답한 경우나 조사원이 잘못 기입한 경우)
- 내용에러에는 코딩에러(코딩에러는 데이터 입력에러를 수반함)와 데이터 입력에러, 내검에러, 결과표 작성에러 등도 있음
- 미상에 대한 처리
  - 데이터 에디팅 팀은 미상을 어떻게 처리할 것인지 센서스 수행 전에 결정해야함
- 미상에 대한 처리가 없다면 많은 이용자들은 미상을 결과표에 같은 비율로 배분하게 될 것임
- 대부분 국가의 기획자들은 미상을 다른 값으로 대체하는 것을 선호함

## □ 무응답대체

- 무응답대체는 유효하지 않는 값과 비일관적인 값에 대한 해법이 될 수 있음 대체는 완전한 값이 에디팅되고 난 후에 처리하는 것이 바람직함
- ICR 방식을 이용한 사전 내검
  - ICR 방식을 이용시 사전에 에러를 제거할 수 있음  
(예: 특정한 빈칸이나 숫자, 응답 정보 등을 이용해서 내검이 가능)
  - 나이, 혼인상태, 성별 등의 항목은 출산력과 연관이 있으므로 이를 이용해서 내검을 할 수 있고 교육정도는 특정한 직업과 연결해 볼 수 있음

## □ 결과공표

- 결과공표 스케줄은 사전에 미리 보도되어야 함
- 결과공표 정책은 국가에 따라 다양하지만 중요 항목에 대한 잠정결과는 자료처리가 끝난 후 1주일정도 이내에 발표되어야 함
- 개인정보는 강력히 보호되어야함
- 중요변수에 대한 집계표를 공표하는 것은 의무적임
- 일부 전문가들은 통계목적이 아닌 용도로 마이크로데이터를 이용하는 것은 민감한 문제이며 이는 다음 센서스에서 의 응답거부로 이어질 수 있으므로 유의해야한다고 말함

## 5. 역할 연극

### □ 수행 방식

- 참석자 그룹을 A~D로 나누어 팀별 1인이 주어진 상황에 맞게 역할을 대변하는 형식임
- 사전에 어떤 질문이 나올지, 상대방이 어떤 응답을 할지 알 수 없고 연극과정에서 즉석에서 답해야함

### □ 조사자와 조사표 디자인 담당자

- 월급을 받지 않고 일하는 가족종사자의 경제활동상태는 어떻게 표시해야하는지
- 가끔씩 해당 거처에 들르는 사람은 조사해야하는지
- 한가지 이상의 직업을 가지고 있는 경우 소득은 어떻게 기록하는지
- 정신적으로 장애가 있는 경우 어떻게 파악하는지

### □ 응답을 거부하는 가구원과 이를 설득하는 조사원

- 왜 정보를 제공해야하는지 모르겠음
- 인터뷰는 얼마나 걸리는지
- 어떤 질문을 할것인지
- 개인정보는 어떻게 보호되는지

### □ 문제를 발생시키는 응답에 대해 에디팅을 하자는 데이터 분석가와 데이터 처리자

- 성이 같은 커플이 있는 경우 다른 성으로 바뀌어야하는지
- 나이와 학력간의 불일치가 있을때 (12살 아이가 박사학위)
- 직업과 산업간의 불일치 (직업은 변호사인데 산업은 농업인 경우)
- 나이와 가족구성원간의 불일치 (20살 엄마와 10살 아들)
- 직업과 학력간의 불일치 (고졸인 교수)

### □ 조사구를 찾기 어렵다는 조사원과 조사구 설정단계를 설명하고 지도 보는 방법을 알려주는 관리자 등

## IV. 시사점 및 참가소감

### □ 우리나라 인구주택총조사 관련 시사점

2010 센서스의 성공적인 수행을 위해서는 다양한 관점에서의 센서스의 품질관리가 이루어져야함

### □ 체계적인 품질관리 체크리스트 작성 필요

- 우리나라는 누락률이 매우 낮고 (2%대) 센서스 수행이 과정별로 체계화되어 있음
- 그러나 각 단계별로 품질을 관리하고 오류가능성을 가늠할 수 있는 체계적인 체크리스트가 없다는 한계

### □ 품질관리 부문에서 인터넷 강국의 이점 활용

- 한국 센서스 품질관련 프리젠테이션 후 다른 참가국에서 2005 센서스의 사이버교육에 관해 많은 관심을 보임
- 인터넷을 통한 일관된 교육 및 지침전달은 데이터 품질향상에 기여할 수 있음
- 인터넷 조사부문에서 발전국가의 강점을 살려 이를 품질향상에 기여할 수 있는 방안 모색

## □ 참가소감 및 기타사항

- 역할연극을 통해 응답자, 조사원, 항목담당자, 관리자 등 다양한 관점에서 생각해 보고 다양한 상황에서 센서스의 오류가 발생할 수 있다는 점을 체험할 수 있었음
- 회의 참가자들은 각 통계청에서 10여년 이상 근무한 직원들이며 특히 비영어권 국가 참가자들도 영어에 능통하여 토론 및 발표에 능동적으로 참여하는 모습이 인상적이었음
- UNSIAP과 마카오 통계청의 지원으로 숙박, 조식, 중식 무료제공 기타 지원사항을 전담 직원이 처리해주는 등 출장수행에 불편함이 없었음

## ※ 회의 관련 사진

자기소개	오프닝 사진
	
회의장	역할연극
	



## <참고1. 싱가포르의 센서스 개괄>

- 싱가포르는 2000년에 처음 등록센서스를 도입하였고 모든 인구수와 인구에 대한 기본적 사항들은 행정자료로부터 얻었음
- 다만 부가적인 항목에 대해서는 거처단위로 전체의 20%에 대해서는 표본조사를 수행함
  - 혼인상태와 출산력, 교육정도, 집에서 사용하는 언어, 경제활동상태, 교통수단등
- 인구, 가구 주택 관련기본정보를 생산하기 위해서 가구등록데이터베이스 (Household Registration Database)와 거처등록데이터베이스(National Database on Dwellings)를 이용함
  - 기본적인 데이터(성별, 인종, 생년월일, 태어난 국가, 국적, 거주형태, 거처의 형태등)는 위 데이터베이스로부터 얻음
- 센서스 개요
  - 주기 및 센서스 기준일 : 10년주기, 2010년 6월 30일
  - 항목수 : 50개
  - 다루는 주제 : 인구, 사회학적 주제, 국제이동, 가구와 가족항목, 출산력, 경제활동상태, 학력, 고령화와 교통수단에 관한 항목
  - 조사방법 : 인터넷, CATI, 면접방식
  - 무응답률 : 5% 미만으로 추정
  - 데이터 입력 : CATI, 인터넷, PDA 등
  - 무응답대체 : 무응답항목에 대해 조정하지 않음
  - 첫 번째 공식발표 : 기본적인 항목에 대해 센서스 기준일로부터 6개월 이내
  - 마이크로데이터제공 : 하지 않음
  - 센서스 환경변화 : 센서스 수행비용 증가, 프라이버시 보호 증가, 인터넷사용의 증가

## <참고2. 일본의 센서스 개괄>

- 주기 : 매 5년, 0자로 끝나는 해는 대규모, 5자는 소규모 (항목수에서 차이)
- 센서스 기준일 : 10월 1일
- 조사표 : 전수조사표만 있음
- 항목수 : 2000년 22개, 2005 17개
- 조사항목 : 우리와 유사하며 전수항목에 산업, 직업,경제 활동상태, 근무시간, 등이 있음
- 2005년 조사원수 : 9만여명
- 인터넷조사 여부 : 2010 시험조사에서 인터넷조사 가능성을 시험했고 우편조사 mail-back 시스템도 도입 예정, 인터넷조사는 2015년경 확대 예정
- 결과공표 및 결과표 작성
  - 첫 번째 결과표 : 인구, 가구, 주택, 노인관련관련 기본항목만
  - 두 번째 결과표 : 노동력, 산업관련 2자리 코딩
  - 세 번째 : 직업, 여성아동 관련
  - 자세한 표본 결과표 작성  
10~20% 가구를 샘플링해서 산업 직업을 3~4자리로 코딩함
- 마이크로데이터 제공하지 않음