

Data Sources



Victoria Velkoff
U.S. Census Bureau



Data Sources – Outline

Definitions and Data Sources:

- Census
- Sample survey
- Administrative data
- Other sources
- Online data sources
- Advantages
- Limitations
- “Bad” data



Definition: Census

Census –

“the total process of collecting, compiling, evaluating, analyzing, and publishing or otherwise disseminating demographic, economic, and social data pertaining, at a specified time, to all persons in a country or in a well-delimited part of the country.”

Source: United Nations, 1992

In other words, a complete count of the people and houses in a given area for a specific point in time.



Features of a Census

The United Nations lists **four** essential features of a census:

1. Universality
2. Individual enumeration
3. Simultaneity
4. Defined periodicity



Data Collected in a Census

General Demographic

- Age
- Sex
- Relation to household head
- Marital status
- Citizenship
- Ethnicity



Data Collected in a Census, cont.

Fertility and Mortality

- Children ever born
- Children living

Geographic Location

- Place of usual residence
- Place of residence at some specified time in the past



Data Collected in a Census, cont.

Education

- School attendance
- Educational attainment

Labor Force

- Activity status
- Occupation
- Industry



Data Collected in a Census, cont.

Housing and Human Settlement

- Type of building
- Construction material
- Year of construction
- Location of living quarters
- Tenure (rent, own, lease)
- Plumbing
- Etc.



Census – Advantages and Limitations

Advantages –

- Representative of entire country
- Universal – gives detailed data at very small geographic levels
- Wide range of topics covered

Limitations –

- Very expensive
- Long time needed for preparation
- Need a large well trained staff



Census – Limitations, cont.



Limitations –

- Current only every 5 or 10 years
- Range of topics wide, but not detailed
- ‘Hard to find’ populations may not be enumerated completely
- Takes time for data processing
- Often takes years for data to be ready for dissemination



Definition: Sample Survey

Sample Survey –

Sampling techniques are used to select a small proportion of the population that is believed to be representative of the population at large. A survey is then conducted using this sample population to gain estimates for the total population.

Note: the size of the sample will determine if these data are representative of smaller geographic areas.



Sample Survey – Advantages

Advantages –

- Provide current information; data often available within months
- Preparation and processing time is less than for census
- Smaller trained staff than for a census
- More detail on a particular topic
- Less expensive than census



Sample Survey – Limitations

Limitations –

- Sample size determines if results are generalizable to entire population
- Data on small sub-populations may not be reliable
- Information on small geographic areas may not be available
- Limited number of topics covered



Definition: Administrative Data

Administrative data –

Collected by a variety of organizations, such as governments, hospitals, and schools.

Types of administrative data –

- Vital events (births, deaths, marriages)
- Education (enrollment, schools, teachers)
- Health (hospitals, medical staff)
- Employment (employees, enterprises)
- Geographic location (voting registration)

Administrative Data – Advantages and Limitations

Advantages –

- Cover the entire population
- Used to analyze changes and trends
- Strong geographic detail
- Very timely

Limitations –

- Multiple registration → double counting
- Incomplete registration
- Limited content



Other Data Sources

- Demographic surveillance sites
 - Data collected for the entire population in study site
 - Can provide very detailed information of surveillance population
 - Not typically representative to national population
- Business organizations (e.g., organizations of women business owners)
- Other sources.....



Online Data Sources

On-line data sources are becoming more and more useful

● International data sites

- UN sites (UNECE, UNDP, ILO, WHO, UNSD, UNESCAP, UNIFEM, INSTRAW)
- U.S. Census Bureau
- World Bank
- USAID
- Asian Development Bank
- Asia-Pacific Economic Cooperation



Bad Data – how do we know them?

Questions to ask oneself:

1. Who collected the data?
2. How was the data collected?
3. How large is the sample and is the sample nationally representative?
4. Is the wording of the questions clear and easy to understand?
5. Is there incentive for respondents to lie?



Checks on data



Analysts can perform certain checks on published (electronic and printed) data:

- Check for consistency across tables in totals and subtotals
- See that sample size is consistent across variables, to check for missing data
- Look for consistency in data “stories”
- Use professional judgment to see if data makes sense in its context

Data use



- When beginning to work on gender statistics check for data that are already available
- Sometimes data are collected but not published or not published disaggregated by sex
- Request that data be disseminated in a way that is useful to you

Data Sources – Outline

Definitions and Data Sources:

- Census
- Sample survey
- Administrative data
- Other sources
- Online data sources
- Advantages
- Limitations
- “Bad” data

