Identifying Family Relationships for the World’s Largest Collection of Census Microdata

IPUMS-International: Integrated Public Use Microdata Series

Matthew Sobek and Sheela Kennedy

Minnesota Population Center, University of Minnesota

ABSTRACT

To facilitate comparative research on families and households, the IPUMS-International project has developed consistent “pointer” variables identifying each person’s mother, father, and spouse across a census microdata collection spanning 55 countries over five decades. These new variables provide researchers with a common tool for studying family interrelationships, removing the possibility that differing results are artifacts of different linking procedures.

This poster documents the methodology used to identify the most likely parent-child and spouse pairings. We show that the IPUMS pointers agree with direct reports of family interrelationships more than 98% of the time, and highlight for researchers and policymakers the factors that affect the precision of these links.

FAMILY INTERRELATIONSHIPS

Family interrelationships in Census microdata

• Census data are typically organized hierarchically: individuals nested in households.
• Each person’s relationship to a reference person in their household is usually known, but relationships to other persons are generally unknown.
• If these family relationships are identified, researchers can combine the characteristics of related and co-resident persons to create new variables and measures, for example:
  - spouse’s age
  - mother’s educational attainment
  - own-child fertility

What are “pointers”?

• Pointers are indicator variables—variables that identify each person’s mother, father, or spouse, if one is present in the household.
• The person indicator variable identifies the person number of a person’s spouse.
• The mother and father indicator variables “point” to a person’s mother and father.

Example: Family Interrelationships in a Census Household

1. head  46 male  married  n/a  0 0
2. spouse  44 female  married  3 1
3. child  15 female  single  0 0
4. child  13 female  single  n/a  0
5. grandchild  3 male  single  n/a  0

INFERRENTIAL POINTERS

How are family relationships inferred?

• Family interrelationships are identified using relationship to head, marital status, fertility, and proximity in the household roster.
• IPUMS pointers are created simultaneously for all countries, applying the same method despite variation in data quality, availability, and household structure.

Empirical refinements

• Comparison with census pointers (direct reports of spouse or parent location)
• Using samples with detailed relationship data to test rules for less detailed samples

METHODOLOGY

General Design of Pointer Process

• Identify allowable relationship-pairing rules
• Set criteria for links within relationship-pairing
• E.g., age-differences, adjacency, marital status, fertility history
• Stronger criteria employed when relationship categories become more ambiguous
• Sample-specific rules when necessary

IPUMS pointer accuracy

We use census pointers (direct reports of spouse and parent line numbers) to evaluate the final IPUMS pointers for the overall rate of disagreement (see table) and factors affecting accuracy (see conclusions).

ACKNOWLEDGEMENTS

Support for this research was provided by the National Science Foundation and the National Institutes of Child Health and Human Development. We are grateful to Vladimir Vladimirov and Moritz Hindman for invaluable technical and design assistance.