

Creating Family, Household, and Neighborhood Level Data from Individual Records

Hsueh-Sheng Wu
CFDR Workshop Series
March 2, 2015

BGSU



Center for
**Family and
Demographic** Research

Outline of Presentation

- Why do we need to create new variables at the aggregate levels
- What are aggregate-level variables?
- The difficulties of constructing aggregate-level variables
- Examples of constructing family-level variables
 - Sample data
 - The characteristics of family
 - The relations of family members
- The sample Stata and SAS command files
 - The size of a family
 - The number of kids in the family
 - The total family income
 - The ID for the spouse of a respondent
 - The number of biological kids for a father

BGSU



Center for

Family and

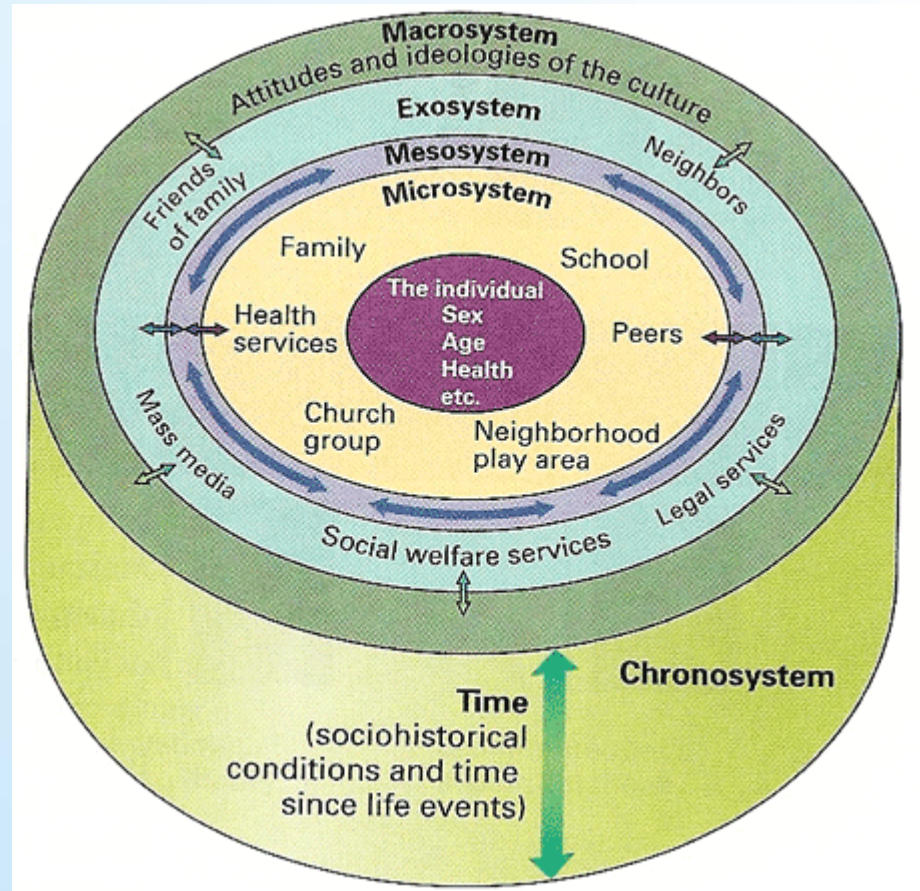
Demographic

Research

- Conclusion

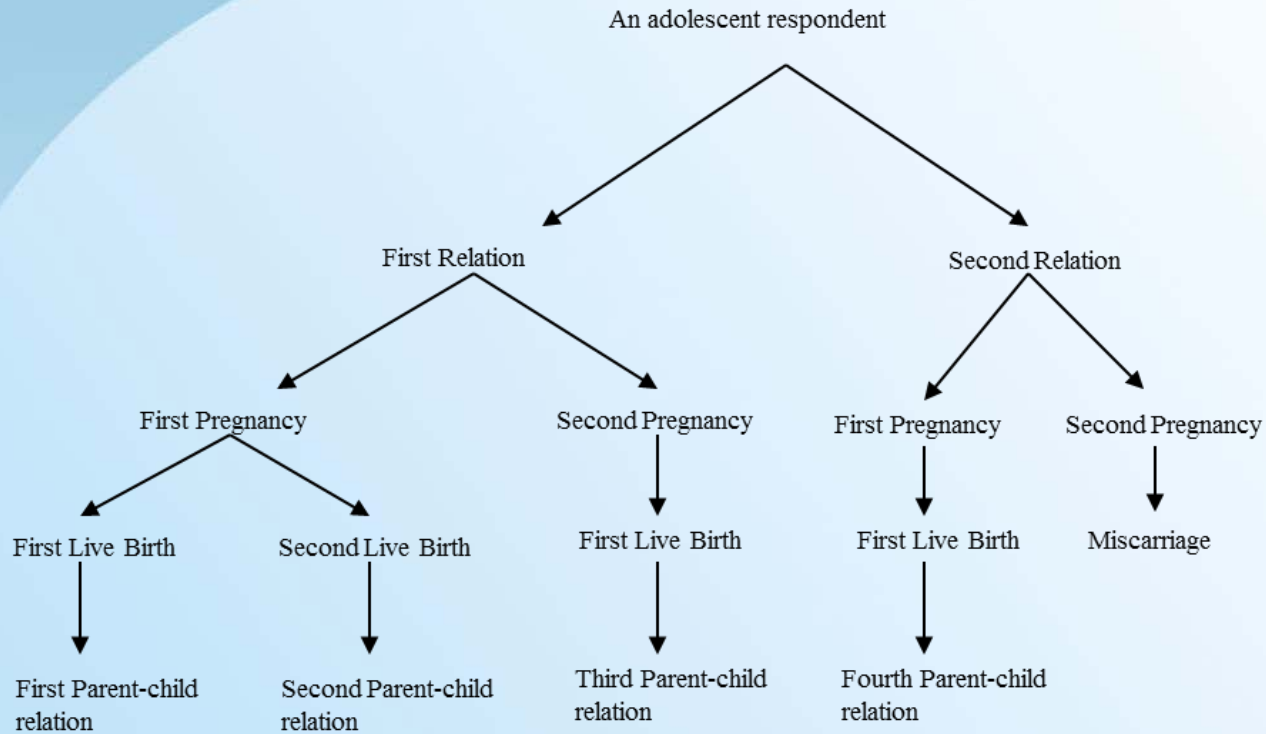
Why do we need aggregate-level variables?

- Examining the relations between family and individuals is the first step toward understanding the interaction between individuals and their environments.
- Once you know how to constructing the family-level variables, you can apply the same methods to construct variables measured at the levels of social networks, schools, communities.



A diagram of Ecological Systems Theory by Bronfenbrenner (Santrock, J. W. (2007).

Why do we need aggregate-level variables? (Continued)



aid	rrelno	rpregno	birthno	aid_kid	aid_rel	aid_preg	rel_preg	rel_kid	preg_kid
1	1	1	1	4	2	4	2	3	2
1	1	1	2	4	2	4	2	3	2
1	1	2	1	4	2	4	2	3	1
1	2	1	1	4	2	4	2	1	1
1	2	2	0	4	2	4	2	1	0

Why do we need aggregated-level variables (Continued)

- The **Microsystem** contains the immediate environments that the child is a part of (family, school, peer group, neighborhood, and childcare environments)
- The **Mesosystem** is comprised of connections between the child's immediate environments (i.e., a child's home and school).
- **Exosystem** contains the external environmental settings and other social systems that do not contain the developing child but indirectly affect development (e.g. a parent's workplace, neighborhoods institutions, the media, the government, the economy etc.).
- **Macrosystem** contains all of the various subsystems and the general beliefs and values of the culture, and is made up of written and unwritten principles that regulate everybody's behavior. These principles- whether legal, economic, political, religious, or educational- endow individual life with meaning and value and control the nature and scope of the interactions between the various levels of the total social system.

What are the family-levels variables?

Family-level variables describes either the characteristics of the whole family or the relations among family members

The characteristics of the family

- The size of a family
- The number of kids in the family
- The total of family income

- The relations among family members
 - the ID of the respondent's spouse?
 - The number of biological kids do fathers have?

BGSU



Center for **Family and**
Demographic Research

The difficulties of constructing aggregate-level variables

- Aggregate-level variables is composite variables from created from multiple variables if the data are in the wide format or from multiple observations if the data are in the long format.
- It is easier to construct the these aggregate-level variables if your data are in the long format than in the wide format. You can use the “reshape” command in Stata to change the format of the data and Stat/transfer to transfer data between SAS and Stata.
- It may takes many steps to construct aggregate-level variables, so remember to check your results at each step.

Examples of constructing family-level variables

- Sample Data

Table 1. The sample data from three families

family_id	epppnum	person_id	age	sex	income	epnspous	epnmom	etypmom	epndad	etypdad
1	101	1001	23	1	3000	102
1	102	1002	23	2	.	101
2	101	1003	29	1	.	102
2	102	1004	32	2	5000	101
2	103	1005	6	1	.	.	102	1	101	1
3	101	1006	51	2	2000	102
3	102	1007	53	1	3000	101
3	103	1008	15	2	.	.	101	1	102	2
3	104	1009	3	1	.	.	101	1	102	1

Conclusions

- Sociologists often are interested in examining how social contexts influence behaviors for individuals. Thus, it is important to learn how to construct family-level variables. You can apply the same methods to construct variables measured at other contextual levels.
- When constructing the family-level variables, you would like the data to be in long format than in wide format. When data are in the long format, you can easily change the order of the observations within the family and replaced the values of variables, which can be difficult to do if data are in wide format.
- At each step of constructing family-level variables, please remember to check the data and make sure that your command files do as they are supposed to do.

- If you have any questions, please contact Hsueh-Sheng Wu @ 372-3119 or wuh@bgsu.edu

