

## Using arrays in hierarchical data

Many times data are organized hierarchically. That means you have several observations per household and/or per family. For instance, in the PUMS there is a household record and a person record. The data look like this-

<u>Household Id</u>	<u>Person Id</u>	<u>Age</u>	<u>Gender</u>	<u>Marstat</u>	<u>POB</u>	<u>Educat</u>	<u>Hhinc</u>	<u>Employ</u>
123	1	67	1	1	10	12	45000	3
123	2	66	2	1	70	12	45000	3
456	1	42	1	1	12	16	75000	1
456	2	38	2	1	12	16	75000	1
456	3	14	1	.	12	.	75000	.
456	4	11	1	.	12	.	75000	.
789	1	36	2	2	46	10	24000	2
789	2	10	2	.	32	.	24000	.
789	3	8	2	.	32	.	24000	.

You can see that there are multiple persons with different ID's (ppid) in the same household (hhid). Some variables such as household income (hhinc) are the same for all members.

It is important to understand the 'first and last' method when manipulating datasets and using arrays. The following example will illustrate how to use first and last.