libname in "f:\family workshop\";

* 1. The size of a family
*****************************************************

table in.family_workshop2 as
select *,
count(epppnum) as family_N /* use the variable, family_N, to indicate the total number of people in family */
from in.family_workshop

group by family_id;
quit;

Proc print data = in.family_workshop2;
var family_id epppnum person_id family_N;
run;

* Question 2: The number of kids in the family
*****************************************************

DATA in.family_workshop2;
set in.family_workshop;

kid =0;
if age < 18 & age ~=. then kid =1; /* Note: the variable, kid, indicates whether a family member is a kid or not */

run;

Proc SQL;
create table in.family_workshop3 as
select *
sum(kid) as t_kid /* Note: the variable, t_kid, indicates the total number of kids within the family */
from in.family_workshop2


group by family_id;
quit;

proc print data = in.family_workshop3;
var family_id epppnum person_id age epnmom etypmom epndad etypdad kid t_kid;
run;

* The total family income
*****************************************************
Proc SQL;
create table in.family_workshop2 as
select *
, sum(income) as t_income /* Note: the variable, t_income, indicates the total income of the family */
from in.family_workshop
group by family_id;
quit;

proc print data = in.family_workshop2;
var family_id epppnum person_id income t_income;
run;

**************************************************************
* 4. The ID of the respondent’s spouse?***********************;
**************************************************************;

DATA in.family_workshop2;
set in.family_workshop;
keep family_id epppnum person_id sex epnspous epnwife epnhus;
 **************************************************************
* Generate the indicator for husbands and wives within the family
**************************************************************;
if sex =1 & epnspous ~=. then epnwife = epnspous ;
if sex =2 & epnspous ~=. then epnhus = epnspous;
run;

**************************************************************
* Expand the indicators to the whole family
**************************************************************;

Proc SQL;
create table in.family_workshop3 as
select *
, max(epnwife) as epnwife2,
max(epnhus) as epnhus2
from in.family_workshop2
group by family_id;
quit;

**************************************************************
* Extract the IDs for husband and wife in the family
**************************************************************;

DATA in.family_workshop4;
set in.family_workshop3;
if sex =2 & epppnum = epnwife2 then id_wife = person_id;
if sex =1 & epppnum = epnhus2 then id_hus = person_id;
run;

*********************************************************
* Expand the IDs for husband and wife to the whole family            
*********************************************************;

Proc SQL;
create table in.family_workshop5 as
select *
  ,max(id_wife) as id_wife2,
  max(id_hus) as id_hus2
from in.family_workshop4
group by family_id;
quit;

*********************************************************
* Remove unnecessary information                  
*********************************************************;

DATA in.family_workshop6;
set in.family_workshop5;
if epnspous =. then  id_wife2 =.;
else if epnspous ~= . & sex =2 then  id_wife2 =.;
if  epnspous =. then  id_hus2 =.;
else if epnspous ~= . & sex =1 then  id_hus2 =.;
run;
proc print data =in.family_workshop6;
var family_id epppnum person_id sex epnspous id_wife2 id_hus2;
run;

**********************************************
* 5. How many biological kids do respondents have?            
**********************************************

*************************************************************************
* Count the possible number of biological kids for a ther in the family      
*************************************************************************

data in.family_workshop2;
set in.family_workshop;
  if etypdad =1 then bio_kid =1;
run;

Proc SQL;
create table in.family_workshop3 as
select *
  ,count(bio_kid) as t_bio_kid
from in.family_workshop2
group by family_id;
quit;

proc print data =in.family_workshop3;
var family_id epppnum person_id sex epndad etypdad bio_kid t_bio_kid;
run;

************************************************************
* Extract the ID of biological father's from the kid's report
************************************************************;

data in.family_workshop4;
set in.family_workshop3;
if bio_kid =1 then bio_dad_epn = epndad;
run;

proc print data =in.family_workshop4;
var family_id epppnum person_id sex epndad etypdad bio_kid t_bio_kid bio_dad_epn;
run;

****************************************************************************
* Expand the ID of biological father's to the whole family
****************************************************************************;

Proc SQL;
create table in.family_workshop5 as
select *,
max(bio_dad_epn) as m_bio_dad_epn
from in.family_workshop4
group by family_id;
quit;

proc print data =in.family_workshop5;
var family_id epppnum person_id sex epndad etypdad bio_kid t_bio_kid bio_dad_epn m_bio_dad_epn;
run;

**************************************************************************
* Create an indicator of whether adult's epppnum corresponds to the
* ID number that children reported for their biological father.
**************************************************************************;

data in.family_workshop6;
set in.family_workshop5;
nbio_kid = 0;
if epppnum = m_bio_dad_epn & sex = 1 then nbio_kid = 1;
run;

Proc print data =in.family_workshop6;
var family_id epppnum person_id sex epndad etypdad bio_kid t_bio_kid bio_dad_epn m_bio_dad_epn nbio_kid;
run;