ASSIGNMENT #3 *** ANSWERS ***

You are given a SAS data set named DEATH99 (available on the class web site) that contains observations for deaths that occurred in 1999 to residents of Albany, Schenectady, Rensselaer, and Saratoga counties.

libname x 'k:\epi514\datasets';

*** don't alter the original data set - make a NEW one
*** part D asks for an age at death
*** add that variable to the data set;

title 'CONTENTS OF ORIGINAL DATA SET: DEATH99';
proc contents data=x.death99;
ods select variables;
run;

data new;
set x.death99;
*calculate age at death;
aad = (dod - dob) / 365;
aad = int((dod - dob) / 365);
format aad 5.1;
label aad = 'AGE AT DEATH';
run;

title "RANDOM OBSERVATIONS FROM DATA SET 'NEW' WITH AAD (AGE AT DEATH) ADDED";
proc print data=new;
where ranuni(0) le .005;
run;
*** part A/ observations for ONLY Albany county;

data albany;
set new;

* subsetting IF;
if res eq 1;

* IF-THEN-DELETE;
* if res ne 1 then delete;

* IF-THEN-OUTPUT;
* if res eq 1 then output;

* IF-THEN-RETURN (why doesn't this work);
* if res ne 1 then return;

* could also use a WHERE statement;
* where res eq 1;

drop res;
run;

run;

title 'RANDOM OBSERVATIONS FROM DATA SET: ALBANY';
proc print data=albany;
where ranuni(0) le .005;
run;

* EXTRA;

title 'RANDOM OBSERVATIONS FROM ALBANY COUNTY IN DATA SET DEATH99';
proc print data=x.death99;
where res eq 1 and ranuni(0) le .005;
run;
*** part B/ breast cancer and prostate cancer;

data ca_breast ca_prostate;
set new;

if cau eq : 'C50' then output ca_breast;
else if cau eq : 'C61' then output ca_prostate;

* or use an IN operator;
* if cau in ( 'C500' 'C501' 'C502' 'C503' 'C504' 'C505' 'C506' 'C507' 'C508' 'C509');

drop cau;
run;

* EXTRA ... IF AND WHERE IN THE SAME DATA STEP;

data ca_breast ca_prostate;
set new;
where cau in : ('C50' 'C61');

* only breast and prostate cancer are available at this point
if the cause is not breast cancer, it must be prostate cancer
;
if cau eq : 'C50' then output ca_breast;
else output ca_prostate;

drop cau;
run;
*** part C/ males and females;

data males females;
set new;

if gen eq '1' then output males;
else               output females;

drop gen;
run;

*** part D/ print observations where deceased pregnant w/n last 6 months;

proc print data = new label;
where prg eq '1';
var prg gen pla cau aad;
run;
* EXTRA ... you can make all data sets in one pass through the data;

    data
    albany   (drop=res)
    breast   (drop=cau)
    prostate (drop=cau)
    males    (drop=gen)
    females  (drop=gen)
    ;
    set x.death99;

    aad = (dod - dob) / 365;

    * check residence;
    if res eq 1 then output albany;

    * check cause of death;
    if cau eq : 'C50' then output breast;
    else
    if cau eq : 'C61' then output prostate;

    * check the gender;
    if gen eq '1' then output males;
    else               output females;

    format aad 5.1;
    label aad = 'AGE AT DEATH';
    run;

* EXTRA ... WHERE data set options, but you cannot drop variables;

    data
    albany   (where=(res eq 1))
    breast   (where=(cau eq : 'C50'))
    prostate (where=(cau eq 'C61'))
    males    (where=(gen eq '1'))
    females  (where=(gen eq '2'))
    ;
    set x.death99;

    aad = (dod - dob) / 365;

    format aad 5.1;
    label aad = 'AGE AT DEATH';
    run;