DATA population; /\*Example of population dataset\*/

input uniqueid gender $ age casecontrol; /\*gender is defined as a categorical variable by the following $\*/

cards;

1 f 26 1

2 m 27 1

3 m 30 0

4 f 25 0

5 m 29 0

6 f 26 1

7 f 28 0

8 f 28 0

9 m 25 1

10 m 26 0

11 f 30 0

12 m 29 0

13 f 27 0

14 m 27 0

15 f 26 1

16 f 25 0

17 f 29 0

18 f 30 0

19 m 26 0

20 m 26 1

21 f 30 0

22 m 25 0

23 m 27 0

24 m 28 1

;

run;

%LET agerange = 5; /\*For this example, we have chosen the age range to be 5. Controls can therefore be up to 5 years younger or older than the case\*/

%LET ratio = 3; /\*We have chosen to match 3 controls for each case\*/

DATA cases controls;

SET population;

IF casecontrol = 1 THEN OUTPUT cases;

ELSE OUTPUT controls;

RUN;

PROC FREQ NOPRINT DATA=cases;

TABLES age\*gender/OUT=caseout;

RUN;

%MACRO sample(v\_age, v\_gender, v\_count);

DATA qualify1;

SET controls;

WHERE (&v\_age-&agerange <= age <= &v\_age+&agerange)

AND

(gender = "&v\_gender");

case\_age = &v\_age;

case\_gender = "&v\_gender";

SEED = RANUNI(0);

PROC SORT;

BY SEED;

DATA qualify2;

SET qualify1 NOBS=totobs;

IF \_N\_ <= &v\_count\*&ratio;

IF &v\_count\*&ratio <= totobs THEN tag = 'yes';

ELSE tag = 'no';

PROC APPEND BASE=matches DATA=qualify2 force;

PROC SORT DATA=qualify2 OUT=temp1 (KEEP=uniqueid);

BY uniqueid;

PROC SORT DATA=controls OUT=temp2;

BY uniqueid;

DATA controls; /\* controls already matched are removed and can not be matched again\*/

MERGE temp1(IN=in1) temp2(IN=in2);

BY uniqueid;

IF in2 AND NOT in1;

%MEND sample;

DATA \_NULL\_;

SET caseout;

CALL EXECUTE ('%sample('||age||','||gender||','||count||')');

RUN;

/\*The next part is for testing if any of the cases have not received the required number of controls\*/

PROC FREQ NOPRINT DATA=matches;

TABLES case\_age\*case\_gender/OUT=con\_out;

PROC SORT DATA = caseout(RENAME=

(age=case\_age gender=case\_gender count=case\_cnt));

BY case\_age case\_gender;

PROC SORT DATA = con\_out (RENAME= (count=con\_cnt));

BY case\_age case\_gender;

DATA final (DROP=percent);

MERGE caseout con\_out;

BY case\_age case\_gender;

con\_need = case\_cnt\*&ratio;

IF con\_cnt = . THEN con\_cnt = 0;

diff = con\_cnt-con\_need;

PROC PRINT DATA = final; /\*creates a table showing characteristics of the cases who have not received enough matches and how many matches they are missing\*/

WHERE diff < 0;

TITLE 'Insufficient Matches';

RUN;

**RESULT**

| **Obs** | **case\_age** | **case\_gender** | **case\_cnt** | **con\_cnt** | **con\_need** | **diff** |
| --- | --- | --- | --- | --- | --- | --- |
| **4** | 27 | m | 1 | 2 | 3 | -1 |
| **5** | 28 | m | 1 | 0 | 3 | -3 |

**LOG**

Please see the file titled ‘Log test code’