```
1          OPTIONS NONOTES NOSTIMER NOSOURCE NOSYNTAXCHECK;
72
73         DATA population; /*Example of population dataset*/
74         input uniqueid gender $ age ethnic $ casecontrol; /*gender and ethnic are defined as categorical variables by the
74       ! following $*/
75         cards;

NOTE: The data set WORK.POPULATION has 24 observations and 5 variables.
NOTE: DATA statement used (Total process time):
      real time           0.01 seconds
      cpu time            0.01 seconds


100        ;
101        run;
102
103        %LET agerange = 5; /*For this example, we have chosen the age range to be 5. Controls can therefore be up to 5 years
103      ! younger or older than the case*/
104        %LET ratio = 3; /*We have chosen to match 3 controls for each case*/
105
106        DATA cases controls;
107           SET population;
108           IF casecontrol = 1 THEN OUTPUT cases;
109           ELSE OUTPUT controls;
110        RUN;

NOTE: There were 24 observations read from the data set WORK.POPULATION.
NOTE: The data set WORK.CASES has 7 observations and 5 variables.
NOTE: The data set WORK.CONTROLS has 17 observations and 5 variables.
NOTE: DATA statement used (Total process time):
      real time           0.00 seconds
      cpu time            0.01 seconds


111
112        PROC FREQ NOPRINT DATA=cases;
113           TABLES age*gender*ethnic/OUT=caseout; /*Since we want ‘ethnic’ to be a matching variable, we need to add it here*/
114        RUN;

NOTE: There were 7 observations read from the data set WORK.CASES.
NOTE: The data set WORK.CASEOUT has 5 observations and 5 variables.
NOTE: PROCEDURE FREQ used (Total process time):
      real time           0.01 seconds
      cpu time            0.01 seconds


115
116        %MACRO sample(v_age, v_gender, v_count,v_ethnic); /*Here, ethnic is also added*/
117
118           DATA qualify1;
119              SET controls;
120            WHERE (&v_age-&agerange <= age <= &v_age+&agerange)
121              AND
122              (gender = "&v_gender")
123              AND
124              (ethnic = "&v_ethnic"); /*You will also need to add your extra variable to this step.*/
125
126           case_age = &v_age;
127           case_gender = "&v_gender";
128           case_ethnic = "&v_ethnic"; /* You will also need to add your extra variable to this step.*/
129
130           SEED = RANUNI(0);
131           PROC SORT;
132              BY SEED;
133
134           DATA qualify2;
135              SET qualify1 NOBS=totobs;
136              IF _N_ <= &v_count*&ratio;
137              IF &v_count*&ratio <= totobs THEN tag = 'yes';
138                 ELSE tag = 'no';
139
140           PROC APPEND BASE=matches DATA=qualify2 force; /*new data set matches will contain the matched controls*/
141
142           PROC SORT DATA=qualify2 OUT=temp1 (KEEP=uniqueid);
143              BY uniqueid;
144
145           PROC SORT DATA=controls OUT=temp2;
146              BY uniqueid;
147
148           DATA controls; /*the dataset controls is updated so that the controls already matched are removed and can not be
148      ! matched again*/
149              MERGE temp1(IN=in1) temp2(IN=in2);
150              BY uniqueid;
151              IF in2 AND NOT in1;
152
153        %MEND sample;
154
155        DATA _NULL_;
156           SET caseout;
157           CALL EXECUTE ('%sample('||age||','||gender||','||count||','||ethnic||')'); /* You will also need to add your extra
157      ! variable to this step.*/
```

```
158        RUN;

NOTE: Numeric values have been converted to character values at the places given by: (Line):(Column).
      157:30   157:53
NOTE: There were 5 observations read from the data set WORK.CASEOUT.
NOTE: DATA statement used (Total process time):
      real time           0.00 seconds
      cpu time            0.01 seconds


NOTE: CALL EXECUTE generated line.
1         + DATA qualify1;       SET controls;     WHERE (25-5 <= age <= 25+5)       AND       (gender = "m")       AND
(ethnic = "1");      case_age = 25;    case_gender = "m";    case_ethnic = "1";      SEED = RANUNI(0);

NOTE: There were 3 observations read from the data set WORK.CONTROLS.
      WHERE (age>=20 and age<=30) and (gender='m') and (ethnic='1');
NOTE: The data set WORK.QUALIFY1 has 3 observations and 9 variables.
NOTE: DATA statement used (Total process time):
      real time           0.00 seconds
      cpu time            0.00 seconds


1         +
PROC SORT;       BY SEED;

NOTE: There were 3 observations read from the data set WORK.QUALIFY1.
NOTE: The data set WORK.QUALIFY1 has 3 observations and 9 variables.
NOTE: PROCEDURE SORT used (Total process time):
      real time           0.00 seconds
      cpu time            0.00 seconds


1         +
DATA
2         + qualify2;       SET qualify1 NOBS=totobs;       IF _N_ <= 1*3;       IF 1*3 <= totobs THEN tag = 'yes';          ELSE
tag = 'no';

NOTE: There were 3 observations read from the data set WORK.QUALIFY1.
NOTE: The data set WORK.QUALIFY2 has 3 observations and 10 variables.
NOTE: DATA statement used (Total process time):
      real time           0.00 seconds
      cpu time            0.02 seconds


2         +
PROC APPEND BASE=matches DATA=qualify2 force;

NOTE: Appending WORK.QUALIFY2 to WORK.MATCHES.
NOTE: BASE data set does not exist. DATA file is being copied to BASE file.
NOTE: There were 3 observations read from the data set WORK.QUALIFY2.
NOTE: The data set WORK.MATCHES has 3 observations and 10 variables.
NOTE: PROCEDURE APPEND used (Total process time):
      real time           0.00 seconds
      cpu time            0.00 seconds


2         +
PROC SORT DATA=qualify2 OUT=temp1 (KEEP=uniqueid);       BY uniqueid;

NOTE: There were 3 observations read from the data set WORK.QUALIFY2.
NOTE: The data set WORK.TEMP1 has 3 observations and 1 variables.
NOTE: PROCEDURE SORT used (Total process time):
      real time           0.00 seconds
      cpu time            0.01 seconds


3         +     PROC SORT DATA=controls OUT=temp2;       BY uniqueid;

NOTE: There were 17 observations read from the data set WORK.CONTROLS.
NOTE: The data set WORK.TEMP2 has 17 observations and 5 variables.
NOTE: PROCEDURE SORT used (Total process time):
      real time           0.00 seconds
      cpu time            0.00 seconds


3         +                                                              DATA controls;        MERGE temp1(IN=in1) temp2(IN=in2);
     BY uniqueid;        IF in2 AND NOT in1;

NOTE: There were 3 observations read from the data set WORK.TEMP1.
NOTE: There were 17 observations read from the data set WORK.TEMP2.
NOTE: The data set WORK.CONTROLS has 14 observations and 5 variables.
NOTE: DATA statement used (Total process time):
      real time           0.01 seconds
      cpu time            0.01 seconds


4         + DATA qualify1;       SET controls;     WHERE (26-5 <= age <= 26+5)       AND       (gender = "f")       AND
(ethnic = "1");      case_age = 26;    case_gender = "f";    case_ethnic = "1";      SEED = RANUNI(0);

NOTE: There were 3 observations read from the data set WORK.CONTROLS.
      WHERE (age>=21 and age<=31) and (gender='f') and (ethnic='1');
NOTE: The data set WORK.QUALIFY1 has 3 observations and 9 variables.
NOTE: DATA statement used (Total process time):
```

```
      real time           0.00 seconds
      cpu time            0.00 seconds


4         +
PROC SORT;       BY SEED;

NOTE: There were 3 observations read from the data set WORK.QUALIFY1.
NOTE: The data set WORK.QUALIFY1 has 3 observations and 9 variables.
NOTE: PROCEDURE SORT used (Total process time):
      real time           0.00 seconds
      cpu time            0.01 seconds


4         +
DATA
5         + qualify2;       SET qualify1 NOBS=totobs;       IF _N_ <= 3*3;       IF 3*3 <= totobs THEN tag = 'yes';          ELSE
tag = 'no';

NOTE: There were 3 observations read from the data set WORK.QUALIFY1.
NOTE: The data set WORK.QUALIFY2 has 3 observations and 10 variables.
NOTE: DATA statement used (Total process time):
      real time           0.00 seconds
      cpu time            0.01 seconds


5         +
PROC APPEND BASE=matches DATA=qualify2 force;

NOTE: Appending WORK.QUALIFY2 to WORK.MATCHES.
NOTE: There were 3 observations read from the data set WORK.QUALIFY2.
NOTE: 3 observations added.
NOTE: The data set WORK.MATCHES has 6 observations and 10 variables.
NOTE: PROCEDURE APPEND used (Total process time):
      real time           0.00 seconds
      cpu time            0.00 seconds


5         +
PROC SORT DATA=qualify2 OUT=temp1 (KEEP=uniqueid);       BY uniqueid;

NOTE: There were 3 observations read from the data set WORK.QUALIFY2.
NOTE: The data set WORK.TEMP1 has 3 observations and 1 variables.
NOTE: PROCEDURE SORT used (Total process time):
      real time           0.01 seconds
      cpu time            0.01 seconds


6         +     PROC SORT DATA=controls OUT=temp2;       BY uniqueid;

NOTE: There were 14 observations read from the data set WORK.CONTROLS.
NOTE: The data set WORK.TEMP2 has 14 observations and 5 variables.
NOTE: PROCEDURE SORT used (Total process time):
      real time           0.00 seconds
      cpu time            0.01 seconds


6         +                                                          DATA controls;        MERGE temp1(IN=in1) temp2(IN=in2);
     BY uniqueid;        IF in2 AND NOT in1;

NOTE: There were 3 observations read from the data set WORK.TEMP1.
NOTE: There were 14 observations read from the data set WORK.TEMP2.
NOTE: The data set WORK.CONTROLS has 11 observations and 5 variables.
NOTE: DATA statement used (Total process time):
      real time           0.01 seconds
      cpu time            0.01 seconds


7         + DATA qualify1;       SET controls;     WHERE (26-5 <= age <= 26+5)       AND       (gender = "m")       AND
(ethnic = "2");      case_age = 26;    case_gender = "m";    case_ethnic = "2";      SEED = RANUNI(0);

NOTE: There were 3 observations read from the data set WORK.CONTROLS.
      WHERE (age>=21 and age<=31) and (gender='m') and (ethnic='2');
NOTE: The data set WORK.QUALIFY1 has 3 observations and 9 variables.
NOTE: DATA statement used (Total process time):
      real time           0.01 seconds
      cpu time            0.00 seconds


7         +
PROC SORT;       BY SEED;

NOTE: There were 3 observations read from the data set WORK.QUALIFY1.
NOTE: The data set WORK.QUALIFY1 has 3 observations and 9 variables.
NOTE: PROCEDURE SORT used (Total process time):
      real time           0.00 seconds
      cpu time            0.00 seconds


7         +
DATA
8         + qualify2;       SET qualify1 NOBS=totobs;       IF _N_ <= 1*3;       IF 1*3 <= totobs THEN tag = 'yes';          ELSE
tag = 'no';
```

```
NOTE: There were 3 observations read from the data set WORK.QUALIFY1.
NOTE: The data set WORK.QUALIFY2 has 3 observations and 10 variables.
NOTE: DATA statement used (Total process time):
      real time           0.00 seconds
      cpu time            0.01 seconds


8         +
PROC APPEND BASE=matches DATA=qualify2 force;

NOTE: Appending WORK.QUALIFY2 to WORK.MATCHES.
NOTE: There were 3 observations read from the data set WORK.QUALIFY2.
NOTE: 3 observations added.
NOTE: The data set WORK.MATCHES has 9 observations and 10 variables.
NOTE: PROCEDURE APPEND used (Total process time):
      real time           0.00 seconds
      cpu time            0.00 seconds


8         +
PROC SORT DATA=qualify2 OUT=temp1 (KEEP=uniqueid);       BY uniqueid;

NOTE: There were 3 observations read from the data set WORK.QUALIFY2.
NOTE: The data set WORK.TEMP1 has 3 observations and 1 variables.
NOTE: PROCEDURE SORT used (Total process time):
      real time           0.00 seconds
      cpu time            0.01 seconds


9         +     PROC SORT DATA=controls OUT=temp2;       BY uniqueid;

NOTE: There were 11 observations read from the data set WORK.CONTROLS.
NOTE: The data set WORK.TEMP2 has 11 observations and 5 variables.
NOTE: PROCEDURE SORT used (Total process time):
      real time           0.00 seconds
      cpu time            0.01 seconds


9         +                                                              DATA controls;        MERGE temp1(IN=in1) temp2(IN=in2);
     BY uniqueid;        IF in2 AND NOT in1;

NOTE: There were 3 observations read from the data set WORK.TEMP1.
NOTE: There were 11 observations read from the data set WORK.TEMP2.
NOTE: The data set WORK.CONTROLS has 8 observations and 5 variables.
NOTE: DATA statement used (Total process time):
      real time           0.00 seconds
      cpu time            0.02 seconds


10        + DATA qualify1;       SET controls;     WHERE (27-5 <= age <= 27+5)       AND       (gender = "m")       AND
(ethnic = "1");      case_age = 27;    case_gender = "m";    case_ethnic = "1";      SEED = RANUNI(0);

NOTE: There were 0 observations read from the data set WORK.CONTROLS.
      WHERE (age>=22 and age<=32) and (gender='m') and (ethnic='1');
NOTE: The data set WORK.QUALIFY1 has 0 observations and 9 variables.
NOTE: DATA statement used (Total process time):
      real time           0.00 seconds
      cpu time            0.01 seconds


10        +
PROC SORT;       BY SEED;

NOTE: Input data set is empty.
NOTE: The data set WORK.QUALIFY1 has 0 observations and 9 variables.
NOTE: PROCEDURE SORT used (Total process time):
      real time           0.00 seconds
      cpu time            0.00 seconds


10        +
DATA
11        + qualify2;       SET qualify1 NOBS=totobs;       IF _N_ <= 1*3;       IF 1*3 <= totobs THEN tag = 'yes';          ELSE
tag = 'no';

NOTE: There were 0 observations read from the data set WORK.QUALIFY1.
NOTE: The data set WORK.QUALIFY2 has 0 observations and 10 variables.
NOTE: DATA statement used (Total process time):
      real time           0.00 seconds
      cpu time            0.01 seconds


11        +
PROC APPEND BASE=matches DATA=qualify2 force;

NOTE: Appending WORK.QUALIFY2 to WORK.MATCHES.
NOTE: There were 0 observations read from the data set WORK.QUALIFY2.
NOTE: 0 observations added.
NOTE: The data set WORK.MATCHES has 9 observations and 10 variables.
NOTE: PROCEDURE APPEND used (Total process time):
      real time           0.00 seconds
      cpu time            0.00 seconds
```

```
11        +
PROC SORT DATA=qualify2 OUT=temp1 (KEEP=uniqueid);       BY uniqueid;

NOTE: Input data set is empty.
NOTE: The data set WORK.TEMP1 has 0 observations and 1 variables.
NOTE: PROCEDURE SORT used (Total process time):
      real time           0.00 seconds
      cpu time            0.02 seconds


12        +     PROC SORT DATA=controls OUT=temp2;       BY uniqueid;

NOTE: There were 8 observations read from the data set WORK.CONTROLS.
NOTE: The data set WORK.TEMP2 has 8 observations and 5 variables.
NOTE: PROCEDURE SORT used (Total process time):
      real time           0.00 seconds
      cpu time            0.00 seconds


12        +                                                    DATA controls;        MERGE temp1(IN=in1) temp2(IN=in2);
     BY uniqueid;        IF in2 AND NOT in1;

NOTE: There were 0 observations read from the data set WORK.TEMP1.
NOTE: There were 8 observations read from the data set WORK.TEMP2.
NOTE: The data set WORK.CONTROLS has 8 observations and 5 variables.
NOTE: DATA statement used (Total process time):
      real time           0.00 seconds
      cpu time            0.01 seconds


13        + DATA qualify1;       SET controls;     WHERE (28-5 <= age <= 28+5)       AND       (gender = "m")       AND
(ethnic = "1");      case_age = 28;    case_gender = "m";    case_ethnic = "1";      SEED = RANUNI(0);

NOTE: There were 0 observations read from the data set WORK.CONTROLS.
      WHERE (age>=23 and age<=33) and (gender='m') and (ethnic='1');
NOTE: The data set WORK.QUALIFY1 has 0 observations and 9 variables.
NOTE: DATA statement used (Total process time):
      real time           0.00 seconds
      cpu time            0.00 seconds


13        +
PROC SORT;       BY SEED;

NOTE: Input data set is empty.
NOTE: The data set WORK.QUALIFY1 has 0 observations and 9 variables.
NOTE: PROCEDURE SORT used (Total process time):
      real time           0.00 seconds
      cpu time            0.01 seconds


13        +
DATA
14        + qualify2;       SET qualify1 NOBS=totobs;       IF _N_ <= 1*3;       IF 1*3 <= totobs THEN tag = 'yes';          ELSE
tag = 'no';

NOTE: There were 0 observations read from the data set WORK.QUALIFY1.
NOTE: The data set WORK.QUALIFY2 has 0 observations and 10 variables.
NOTE: DATA statement used (Total process time):
      real time           0.00 seconds
      cpu time            0.01 seconds


14        +
PROC APPEND BASE=matches DATA=qualify2 force;

NOTE: Appending WORK.QUALIFY2 to WORK.MATCHES.
NOTE: There were 0 observations read from the data set WORK.QUALIFY2.
NOTE: 0 observations added.
NOTE: The data set WORK.MATCHES has 9 observations and 10 variables.
NOTE: PROCEDURE APPEND used (Total process time):
      real time           0.00 seconds
      cpu time            0.00 seconds


14        +
PROC SORT DATA=qualify2 OUT=temp1 (KEEP=uniqueid);       BY uniqueid;

NOTE: Input data set is empty.
NOTE: The data set WORK.TEMP1 has 0 observations and 1 variables.
NOTE: PROCEDURE SORT used (Total process time):
      real time           0.00 seconds
      cpu time            0.00 seconds


15        +     PROC SORT DATA=controls OUT=temp2;       BY uniqueid;

NOTE: There were 8 observations read from the data set WORK.CONTROLS.
NOTE: The data set WORK.TEMP2 has 8 observations and 5 variables.
NOTE: PROCEDURE SORT used (Total process time):
      real time           0.00 seconds
```

```
      cpu time            0.00 seconds


15         +                                                                DATA controls;        MERGE temp1(IN=in1) temp2(IN=in2);
     BY uniqueid;        IF in2 AND NOT in1;
159
160        /*The next part is for testing if any of the cases have not received the wanted amount of controls*/
161

NOTE: There were 0 observations read from the data set WORK.TEMP1.
NOTE: There were 8 observations read from the data set WORK.TEMP2.
NOTE: The data set WORK.CONTROLS has 8 observations and 5 variables.
NOTE: DATA statement used (Total process time):
      real time           0.00 seconds
      cpu time            0.01 seconds


162        PROC FREQ NOPRINT DATA=matches;
163           TABLES case_age*case_gender*case_ethnic/OUT=con_out; /* You will also need to add your extra variable to this step.*/
164

NOTE: There were 9 observations read from the data set WORK.MATCHES.
NOTE: The data set WORK.CON_OUT has 3 observations and 5 variables.
NOTE: PROCEDURE FREQ used (Total process time):
      real time           0.01 seconds
      cpu time            0.02 seconds


165        PROC SORT DATA = caseout(RENAME=
166                          (age=case_age gender=case_gender count=case_cnt ethnic=case_ethnic)); /* You will also need to add your
166      ! extra variable to this step.*/
167           BY case_age case_gender case_ethnic; /* You will also need to add your extra variable to this step.*/
168

NOTE: There were 5 observations read from the data set WORK.CASEOUT.
NOTE: The data set WORK.CASEOUT has 5 observations and 5 variables.
NOTE: PROCEDURE SORT used (Total process time):
      real time           0.00 seconds
      cpu time            0.00 seconds


169        PROC SORT DATA = con_out (RENAME= (count=con_cnt));
170           BY case_age case_gender case_ethnic; /* You will also need to add your extra variable to this step.*/
171

NOTE: There were 3 observations read from the data set WORK.CON_OUT.
NOTE: The data set WORK.CON_OUT has 3 observations and 5 variables.
NOTE: PROCEDURE SORT used (Total process time):
      real time           0.00 seconds
      cpu time            0.01 seconds


172        DATA final (DROP=percent);
173           MERGE caseout con_out;
174           BY case_age case_gender case_ethnic; /* You will also need to add your extra variable to this step.*/
175
176        con_need = case_cnt*&ratio;
177        IF con_cnt = . THEN con_cnt = 0;
178        diff = con_cnt-con_need;
179

NOTE: There were 5 observations read from the data set WORK.CASEOUT.
NOTE: There were 3 observations read from the data set WORK.CON_OUT.
NOTE: The data set WORK.FINAL has 5 observations and 7 variables.
NOTE: DATA statement used (Total process time):
      real time           0.00 seconds
      cpu time            0.01 seconds


180        PROC PRINT DATA = final; /*creates a table showing what characterizes the cases who have not received enough matches and
180      ! how many matches they are missing*/
181           WHERE diff < 0;
182           TITLE 'Insufficient Matches';
183        RUN;

NOTE: There were 3 observations read from the data set WORK.FINAL.
      WHERE diff<0;
NOTE: PROCEDURE PRINT used (Total process time):
      real time           0.10 seconds
      cpu time            0.10 seconds


184
185
186        OPTIONS NONOTES NOSTIMER NOSOURCE NOSYNTAXCHECK;
198
```