



Copyright 1985-2017 StataCorp LLC
StataCorp
4905 Lakeway Drive
College Station, Texas 77845 USA
800-STATATA-PC <http://www.stata.com>
979-696-4600 stata@stata.com
979-696-4601 (fax)

25-user Stata network perpetual license:
Serial number: 401506300767
Licensed to: DIW
DIW Berlin

Notes:

1. Unicode is supported; see [help unicode advice](#).
2. Maximum number of variables is set to 5000; see [help set maxvar](#).

running C:\stata15-64\sysprofile.do ...
r; t=0.01 14:43:34

running c:\ado\personal\profile.do ...
r; t=0.02 14:43:34

```
1 . doedit "C:\Users\mwronsk\OneDrive - Szkoła Główna Handlowa w Warszawie\Bank Światowy Rumunia\
> on\Code\1. Imputation of tax income into EU-SILC.do"
r; t=0.08 14:43:35

2 . do "C:\Users\mwronsk\OneDrive - Szkoła Główna Handlowa w Warszawie\Bank Światowy Rumunia\Repl
> ode\1. Imputation of tax income into EU-SILC.do"

3 . **Change to your dictionary**
4 . cd "C:\Users\mwronsk\OneDrive - Szkoła Główna Handlowa w Warszawie\Bank Światowy Rumunia\Repl
C:\Users\mwronsk\OneDrive - Szkoła Główna Handlowa w Warszawie\Bank Światowy Rumunia\Replicatio
r; t=0.00 14:44:02

5 .
6 . **EU-SILC impute, you need to change the file adress. The file is uploaded into 03.modified da
7 . use Data\ROU_SILC_2021, clear
r; t=0.02 14:44:02

8 . encode db040, gen(db040_encode)
r; t=0.01 14:44:02

9 . append using Data\Tax_07_2020_for_imputation
(note: variable rb080 was int, now float to accommodate using data's values)
(note: variable db040_encode was long, now double to accommodate using data's values)
(note: variable nace was long, now double to accommodate using data's values)
(note: variable rb090 was byte, now long to accommodate using data's values)
(note: variable pl145 was byte, now float to accommodate using data's values)
r; t=1.80 14:44:03

10 .
11 .
12 .
13 . global matchingvars rb080 rb090 db040_encode nace pl145
r; t=0.00 14:44:03
```

```

14 . mi set mlong
    r; t=0.45 14:44:04

15 . mi register imputed taxincome
    (16630 m=0 obs. now marked as incomplete)
    r; t=7.03 14:44:11

16 . mi register imputed taxincome_top_coded
    r; t=3.90 14:44:15

17 .
18 . *The imputation takes long time - our tax file is large*
19 . mi impute pmm taxincome ${matchingvars} if p1145!=., knn(10) add(10) rseed(12345) replace

```

```

Univariate imputation                Imputations =      10
Predictive mean matching              added =      10
Imputed: m=1 through m=10             updated =       0

                                   Nearest neighbors =      10

```

Variable	Observations per m			Total
	Complete	Incomplete	Imputed	
taxincome	5994961	6894	6894	6001855

```

(complete + incomplete = total; imputed is the minimum across m
of the number of filled-in observations.)
r; t=24774.18 21:37:09

```

```

20 . keep if taxdata==.
    (5,994,961 observations deleted)
    r; t=0.59 21:37:10

21 . gen original=.
    (182,930 missing values generated)
    r; t=0.01 21:37:10

22 . replace original=1 if _mi_id<16630
    (182,919 real changes made)
    r; t=0.02 21:37:10

23 .
24 . *Variable creation*
25 .
26 . *Data analysis on the imputed tax file*
27 . mi xeq: gen age=2020-pb140

```

```

m=0 data:
-> gen age=2020-pb140
    (1,342 missing values generated)

```

```

m=1 data:
-> gen age=2020-pb140
    (1,342 missing values generated)

```

```

m=2 data:
-> gen age=2020-pb140
    (1,342 missing values generated)

```

```

m=3 data:
-> gen age=2020-pb140
    (1,342 missing values generated)

```

```

m=4 data:
-> gen age=2020-pb140
    (1,342 missing values generated)

```

```
m=5 data:
-> gen age=2020-pb140
(1,342 missing values generated)
```

```
m=6 data:
-> gen age=2020-pb140
(1,342 missing values generated)
```

```
m=7 data:
-> gen age=2020-pb140
(1,342 missing values generated)
```

```
m=8 data:
-> gen age=2020-pb140
(1,342 missing values generated)
```

```
m=9 data:
-> gen age=2020-pb140
(1,342 missing values generated)
```

```
m=10 data:
-> gen age=2020-pb140
(1,342 missing values generated)
r; t=5.16 21:37:15
```

28 . mi xeq: gen compliance_ratio=taxincome/py010g

```
m=0 data:
-> gen compliance_ratio=taxincome/py010g
(16,630 missing values generated)
```

```
m=1 data:
-> gen compliance_ratio=taxincome/py010g
(10,949 missing values generated)
```

```
m=2 data:
-> gen compliance_ratio=taxincome/py010g
(10,949 missing values generated)
```

```
m=3 data:
-> gen compliance_ratio=taxincome/py010g
(10,949 missing values generated)
```

```
m=4 data:
-> gen compliance_ratio=taxincome/py010g
(10,949 missing values generated)
```

```
m=5 data:
-> gen compliance_ratio=taxincome/py010g
(10,949 missing values generated)
```

```
m=6 data:
-> gen compliance_ratio=taxincome/py010g
(10,949 missing values generated)
```

```
m=7 data:
-> gen compliance_ratio=taxincome/py010g
(10,949 missing values generated)
```

```
m=8 data:
-> gen compliance_ratio=taxincome/py010g
(10,949 missing values generated)
```

```
m=9 data:
-> gen compliance_ratio=taxincome/py010g
(10,949 missing values generated)
```

```
m=10 data:
-> gen compliance_ratio=taxincome/py010g
(10,949 missing values generated)
r; t=4.16 21:37:19
```

```
29 . mi xeq: gen age_group=1 if age<25

m=0 data:
-> gen age_group=1 if age<25
(15,227 missing values generated)

m=1 data:
-> gen age_group=1 if age<25
(15,227 missing values generated)

m=2 data:
-> gen age_group=1 if age<25
(15,227 missing values generated)

m=3 data:
-> gen age_group=1 if age<25
(15,227 missing values generated)

m=4 data:
-> gen age_group=1 if age<25
(15,227 missing values generated)

m=5 data:
-> gen age_group=1 if age<25
(15,227 missing values generated)

m=6 data:
-> gen age_group=1 if age<25
(15,227 missing values generated)

m=7 data:
-> gen age_group=1 if age<25
(15,227 missing values generated)

m=8 data:
-> gen age_group=1 if age<25
(15,227 missing values generated)

m=9 data:
-> gen age_group=1 if age<25
(15,227 missing values generated)

m=10 data:
-> gen age_group=1 if age<25
(15,227 missing values generated)
r; t=3.97 21:37:23

30 . mi xeq: replace age_group=2 if age>24 & age<30

m=0 data:
-> replace age_group=2 if age>24 & age<30
(623 real changes made)

m=1 data:
-> replace age_group=2 if age>24 & age<30
(623 real changes made)

m=2 data:
-> replace age_group=2 if age>24 & age<30
(623 real changes made)

m=3 data:
-> replace age_group=2 if age>24 & age<30
(623 real changes made)

m=4 data:
-> replace age_group=2 if age>24 & age<30
(623 real changes made)
```

```
m=5 data:
-> replace age_group=2 if age>24 & age<30
(623 real changes made)
```

```
m=6 data:
-> replace age_group=2 if age>24 & age<30
(623 real changes made)
```

```
m=7 data:
-> replace age_group=2 if age>24 & age<30
(623 real changes made)
```

```
m=8 data:
-> replace age_group=2 if age>24 & age<30
(623 real changes made)
```

```
m=9 data:
-> replace age_group=2 if age>24 & age<30
(623 real changes made)
```

```
m=10 data:
-> replace age_group=2 if age>24 & age<30
(623 real changes made)
r; t=4.03 21:37:28
```

```
31 . mi xeq: replace age_group=3 if age>29 & age<35
```

```
m=0 data:
-> replace age_group=3 if age>29 & age<35
(842 real changes made)
```

```
m=1 data:
-> replace age_group=3 if age>29 & age<35
(842 real changes made)
```

```
m=2 data:
-> replace age_group=3 if age>29 & age<35
(842 real changes made)
```

```
m=3 data:
-> replace age_group=3 if age>29 & age<35
(842 real changes made)
```

```
m=4 data:
-> replace age_group=3 if age>29 & age<35
(842 real changes made)
```

```
m=5 data:
-> replace age_group=3 if age>29 & age<35
(842 real changes made)
```

```
m=6 data:
-> replace age_group=3 if age>29 & age<35
(842 real changes made)
```

```
m=7 data:
-> replace age_group=3 if age>29 & age<35
(842 real changes made)
```

```
m=8 data:
-> replace age_group=3 if age>29 & age<35
(842 real changes made)
```

```
m=9 data:
-> replace age_group=3 if age>29 & age<35
(842 real changes made)
```

```
m=10 data:
-> replace age_group=3 if age>29 & age<35
(842 real changes made)
r; t=4.26 21:37:32
```

```
32 . mi xeq: replace age_group=4 if age>34 & age<40
```

```
m=0 data:
-> replace age_group=4 if age>34 & age<40
(894 real changes made)
```

```
m=1 data:
-> replace age_group=4 if age>34 & age<40
(894 real changes made)
```

```
m=2 data:
-> replace age_group=4 if age>34 & age<40
(894 real changes made)
```

```
m=3 data:
-> replace age_group=4 if age>34 & age<40
(894 real changes made)
```

```
m=4 data:
-> replace age_group=4 if age>34 & age<40
(894 real changes made)
```

```
m=5 data:
-> replace age_group=4 if age>34 & age<40
(894 real changes made)
```

```
m=6 data:
-> replace age_group=4 if age>34 & age<40
(894 real changes made)
```

```
m=7 data:
-> replace age_group=4 if age>34 & age<40
(894 real changes made)
```

```
m=8 data:
-> replace age_group=4 if age>34 & age<40
(894 real changes made)
```

```
m=9 data:
-> replace age_group=4 if age>34 & age<40
(894 real changes made)
```

```
m=10 data:
-> replace age_group=4 if age>34 & age<40
(894 real changes made)
r; t=3.98 21:37:36
```

```
33 . mi xeq: replace age_group=5 if age>39 & age<45
```

```
m=0 data:
-> replace age_group=5 if age>39 & age<45
(1,323 real changes made)
```

```
m=1 data:
-> replace age_group=5 if age>39 & age<45
(1,323 real changes made)
```

```
m=2 data:
-> replace age_group=5 if age>39 & age<45
(1,323 real changes made)
```

```
m=3 data:
-> replace age_group=5 if age>39 & age<45
(1,323 real changes made)
```

```
m=4 data:
-> replace age_group=5 if age>39 & age<45
(1,323 real changes made)
```

```
m=5 data:
-> replace age_group=5 if age>39 & age<45
(1,323 real changes made)
```

```
m=6 data:
-> replace age_group=5 if age>39 & age<45
(1,323 real changes made)
```

```
m=7 data:
-> replace age_group=5 if age>39 & age<45
(1,323 real changes made)
```

```
m=8 data:
-> replace age_group=5 if age>39 & age<45
(1,323 real changes made)
```

```
m=9 data:
-> replace age_group=5 if age>39 & age<45
(1,323 real changes made)
```

```
m=10 data:
-> replace age_group=5 if age>39 & age<45
(1,323 real changes made)
r; t=4.05 21:37:40
```

34 . mi xeq: replace age_group=6 if age>44 & age<50

```
m=0 data:
-> replace age_group=6 if age>44 & age<50
(1,384 real changes made)
```

```
m=1 data:
-> replace age_group=6 if age>44 & age<50
(1,384 real changes made)
```

```
m=2 data:
-> replace age_group=6 if age>44 & age<50
(1,384 real changes made)
```

```
m=3 data:
-> replace age_group=6 if age>44 & age<50
(1,384 real changes made)
```

```
m=4 data:
-> replace age_group=6 if age>44 & age<50
(1,384 real changes made)
```

```
m=5 data:
-> replace age_group=6 if age>44 & age<50
(1,384 real changes made)
```

```
m=6 data:
-> replace age_group=6 if age>44 & age<50
(1,384 real changes made)
```

```
m=7 data:
-> replace age_group=6 if age>44 & age<50
(1,384 real changes made)
```

```
m=8 data:
-> replace age_group=6 if age>44 & age<50
(1,384 real changes made)
```

```
m=9 data:
-> replace age_group=6 if age>44 & age<50
(1,384 real changes made)
```

```
m=10 data:
-> replace age_group=6 if age>44 & age<50
(1,384 real changes made)
r; t=4.00 21:37:44
```

35 . mi xeq: replace age_group=7 if age>49 & age<55

m=0 data:
-> **replace age_group=7 if age>49 & age<55**
(1,557 real changes made)

m=1 data:
-> **replace age_group=7 if age>49 & age<55**
(1,557 real changes made)

m=2 data:
-> **replace age_group=7 if age>49 & age<55**
(1,557 real changes made)

m=3 data:
-> **replace age_group=7 if age>49 & age<55**
(1,557 real changes made)

m=4 data:
-> **replace age_group=7 if age>49 & age<55**
(1,557 real changes made)

m=5 data:
-> **replace age_group=7 if age>49 & age<55**
(1,557 real changes made)

m=6 data:
-> **replace age_group=7 if age>49 & age<55**
(1,557 real changes made)

m=7 data:
-> **replace age_group=7 if age>49 & age<55**
(1,557 real changes made)

m=8 data:
-> **replace age_group=7 if age>49 & age<55**
(1,557 real changes made)

m=9 data:
-> **replace age_group=7 if age>49 & age<55**
(1,557 real changes made)

m=10 data:
-> **replace age_group=7 if age>49 & age<55**
(1,557 real changes made)
r; t=4.03 21:37:48

36 . mi xeq: replace age_group=8 if age>54 & age<60

m=0 data:
-> **replace age_group=8 if age>54 & age<60**
(1,015 real changes made)

m=1 data:
-> **replace age_group=8 if age>54 & age<60**
(1,015 real changes made)

m=2 data:
-> **replace age_group=8 if age>54 & age<60**
(1,015 real changes made)

m=3 data:
-> **replace age_group=8 if age>54 & age<60**
(1,015 real changes made)

m=4 data:
-> **replace age_group=8 if age>54 & age<60**
(1,015 real changes made)


```
m=5 data:
-> replace age_group=8 if age>54 & age<60
(1,015 real changes made)
```

```
m=6 data:
-> replace age_group=8 if age>54 & age<60
(1,015 real changes made)
```

```
m=7 data:
-> replace age_group=8 if age>54 & age<60
(1,015 real changes made)
```

```
m=8 data:
-> replace age_group=8 if age>54 & age<60
(1,015 real changes made)
```

```
m=9 data:
-> replace age_group=8 if age>54 & age<60
(1,015 real changes made)
```

```
m=10 data:
-> replace age_group=8 if age>54 & age<60
(1,015 real changes made)
r; t=3.98 21:37:52
```

37 . mi xeq: replace age_group=9 if age>59 & age<65

```
m=0 data:
-> replace age_group=9 if age>59 & age<65
(1,352 real changes made)
```

```
m=1 data:
-> replace age_group=9 if age>59 & age<65
(1,352 real changes made)
```

```
m=2 data:
-> replace age_group=9 if age>59 & age<65
(1,352 real changes made)
```

```
m=3 data:
-> replace age_group=9 if age>59 & age<65
(1,352 real changes made)
```

```
m=4 data:
-> replace age_group=9 if age>59 & age<65
(1,352 real changes made)
```

```
m=5 data:
-> replace age_group=9 if age>59 & age<65
(1,352 real changes made)
```

```
m=6 data:
-> replace age_group=9 if age>59 & age<65
(1,352 real changes made)
```

```
m=7 data:
-> replace age_group=9 if age>59 & age<65
(1,352 real changes made)
```

```
m=8 data:
-> replace age_group=9 if age>59 & age<65
(1,352 real changes made)
```

```
m=9 data:
-> replace age_group=9 if age>59 & age<65
(1,352 real changes made)
```

```
m=10 data:
-> replace age_group=9 if age>59 & age<65
(1,352 real changes made)
r; t=4.18 21:37:57
```

```
38 . mi xeq: replace age_group=10 if age>64
```

```
m=0 data:
-> replace age_group=10 if age>64
(6,237 real changes made)
```

```
m=1 data:
-> replace age_group=10 if age>64
(6,237 real changes made)
```

```
m=2 data:
-> replace age_group=10 if age>64
(6,237 real changes made)
```

```
m=3 data:
-> replace age_group=10 if age>64
(6,237 real changes made)
```

```
m=4 data:
-> replace age_group=10 if age>64
(6,237 real changes made)
```

```
m=5 data:
-> replace age_group=10 if age>64
(6,237 real changes made)
```

```
m=6 data:
-> replace age_group=10 if age>64
(6,237 real changes made)
```

```
m=7 data:
-> replace age_group=10 if age>64
(6,237 real changes made)
```

```
m=8 data:
-> replace age_group=10 if age>64
(6,237 real changes made)
```

```
m=9 data:
-> replace age_group=10 if age>64
(6,237 real changes made)
```

```
m=10 data:
-> replace age_group=10 if age>64
(6,237 real changes made)
r; t=3.99 21:38:01
```

```
39 . mean compliance_ratio, over(age_group)
```

```
Mean estimation      Number of obs   =      56,810
```

```
1: age_group = 1
2: age_group = 2
3: age_group = 3
4: age_group = 4
5: age_group = 5
6: age_group = 6
7: age_group = 7
8: age_group = 8
9: age_group = 9
10: age_group = 10
```

Over	Mean	Std. Err.	[95% Conf. Interval]	
compliance_ratio				
1	1.546568	.2304809	1.094824	1.998312
2	.652124	.0359598	.5816426	.7226054
3	.6559071	.0159537	.6246376	.6871765
4	.6071361	.0138643	.5799621	.6343102
5	.6111107	.0130012	.5856283	.636593
6	.8453013	.0660836	.7157772	.9748255
7	.6647255	.0117128	.6417683	.6876827
8	.6542166	.01949	.6160161	.6924171
9	.6351071	.0237068	.5886416	.6815726
10	.8403257	.1341124	.5774645	1.103187

r; t=1.83 21:38:03

```

40 .
41 . save Data\EU_SILC_imputed
file Data\EU_SILC_imputed.dta already exists
r(602); t=0.01 21:38:03

end of do-file

r(602); t=24841.07 21:38:03

42 . save Data\EU_SILC_imputed, replace
file Data\EU_SILC_imputed.dta saved
r; t=0.12 21:44:43

43 .

```