



```

        name: <unnamed>
        log: /hdir/0/jhaber/Projects/charter_data/sorting-schools-2019/logs/multiple_i
> mputation_100_101019.smcl
        log type: smcl
        opened on: 11 Oct 2019, 07:09:59

1 .
2 . * Import prepared data (not yet imputed):
3 . use "data/charter_schools_data_no_imputations.dta", clear

4 .
5 . * Drop vars we don't impute directly
6 . drop povertyschoolcount

7 .
8 . * Make original zeros hard-missing (.z), this way they won't be imputed and can be r
> estored later:
9 . replace lnage = .z if lnage==. & age!=.
   (0 real changes made)

10. replace lnstudents = .z if lnstudents==. & students!=.
    (11 real changes made, 11 to missing)

11. replace lnteachers = .z if lnteachers==. & teachers!=.
    (49 real changes made, 49 to missing)

12.
13.
14. ** -----
15. ** PERFORM MULTIPLE IMPUTATION
16. ** -----
17.
18. * Look at variables missing cases:
19. *mvpatterns inquiry_seed_log inquiry_narrow_log inquiry_full_log inquiry_full_nohand
> s_log povertyschool pocschool povertysd pocsd primary middle high otherlevel age stu
> dents urban cmo titlei readall13 readall14 readall15 mathall13 mathall14 mathall15 e
> xpulsions suspensions incidents lawreferrals lepratio disabledratio approgram teache
> rs teacherratio certcount certrate

20.
21. /* List of vars to impute, both those used in models and those that help with multip
> le imputation:
> * School students and controls:
> pocschoolcount lepcount disabledcount students primary middle high otherlevel age ur
> ban titlei cmo
> * School quality/discipline:
> readall13 readall14 readall15 mathall13 mathall14 mathall15 teachers certcount suspe
> nsions expulsions incidents lawreferrals
> * School district level (add publicdensity charterdensity ??)
> povertysd pocsd childpovertysd collegesd popdensity unemployment foreignborn closera
> te publicdensity
> */

22.
23. * Declare MI data using memory-efficient style best-suited to modifying variables (r
> ather than cases):
24. mi set mlong

25.
26. * Register as imputed those variables missing any cases:
27. mi register imputed lnage lnstudents lnteachers readall13 readall14 readall15 mathal
> l13 mathall14 mathall15 povertyschool povertysd pocsd readlevel13 readlevel14 readle
> vel15 mathlevel13 mathlevel14 mathlevel15
   (2608 m=0 obs. now marked as incomplete)

```

```

28.
29. * Register as regular those vars with complete observations (may be used to help in
> imputation equations):
30. mi register regular closerate publicdensity charterdensity urban titlei ///
> cmo cmonum state statenum geodistrict numwords numpages numpdfs ///
> inquiry_seed_count inquiry_seed_prop inquiry_seed_log inquiry_narrow_count inquiry_n
> arrow_prop inquiry_narrow_log ///
> inquiry_full_count inquiry_full_prop inquiry_full_log inquiry_full_nohands_count inq
> uiry_full_nohands_prop inquiry_full_nohands_log ///
> traditional_prop progressive_prop ///
> pocschool pocschoolcount ///
> ethnicisolated99 ethnicisolated95 ethnicisolated90 ethnicisolated80 ethnicisolated70

31.
32. mi register passive pctpdfs

33.
34. * For reproducibility, set random seed:
35. set seed 43

36. * Increase maximum variable limit:
37. set matsize 5000

38.
39. * Execute multiple implementation:
40. mi impute chained ///
> ///
> (pmm, knn(5) omit(i.readlevel13 i.readlevel14 i.readlevel15 i.mathlevel13 i.mathleve
> l14 i.mathlevel15 lstudents lteachers readall13 readall14 readall15 mathall13 math
> all14 mathall15)) ///
> povertyschool ///
> ///
> (pmm, knn(5) omit(lstudents i.readlevel13 i.readlevel14 i.readlevel15 i.mathlevel13
> i.mathlevel14 i.mathlevel15 mathall13 mathall14 mathall15)) ///
> povertysd ///
> ///
> (pmm, knn(5) omit(i.readlevel13 i.readlevel14 i.readlevel15 i.mathlevel13 i.mathleve
> l14 i.mathlevel15 lstudents readall13 readall14 readall15 mathall13 mathall14 matha
> ll15 lteachers)) ///
> pocsd ///
> ///
> (pmm, knn(5)) lnage ///
> ///
> (pmm, knn(5) omit(lstudents lteachers)) ///
> readall13 ///
> ///
> (pmm, knn(5) omit(lnage povertysd pocsd lstudents readall13 readall14 readall15 lnt
> eachers)) ///
> mathall13 ///
> ///
> (pmm, knn(5) omit(lstudents lteachers)) ///
> readall14 ///
> ///
> (pmm, knn(5) omit(lnage povertysd pocsd lstudents readall13 readall14 readall15 lnt
> eachers)) ///
> mathall14 ///
> ///
> (pmm, knn(5) omit(lstudents lteachers)) ///
> readall15 ///
> ///
> (pmm, knn(5) omit(lnage povertysd pocsd lstudents readall13 readall14 readall15 lnt
> eachers)) ///
> mathall15 ///
> ///
> (pmm, knn(5) omit(lnage povertysd pocsd readall13 readall14 readall15 mathall13 math
> all14 mathall15 lteachers i.readlevel13 i.readlevel14 i.readlevel15 i.mathlevel13 i
> .mathlevel14 i.mathlevel15)) ///
> lstudents ///
> ///
> (pmm, knn(5) omit(lnage povertysd pocsd lstudents readall13 readall14 readall15)) /
> //
> lteachers ///

```

```

> ///
> (ologit, omit(lnage povertysd pocsd lnstudents readall13 readall14 readall15 mathall
> 13 mathall14 mathall15 lnteachers)) ///
> readlevel13 ///
> ///
> (ologit, omit(lnage povertysd pocsd lnstudents readall13 readall14 readall15 mathall
> 13 mathall14 mathall15 lnteachers)) ///
> mathlevel13 ///
> ///
> (ologit, omit(lnage povertysd pocsd lnstudents readall13 readall14 readall15 mathall
> 13 mathall14 mathall15 lnteachers)) ///
> readlevel14 ///
> ///
> (ologit, omit(lnage povertysd pocsd lnstudents readall13 readall14 readall15 mathall
> 13 mathall14 mathall15 lnteachers)) ///
> mathlevel14 ///
> ///
> (ologit, omit(lnage povertysd pocsd lnstudents readall13 readall14 readall15 mathall
> 13 mathall14 mathall15 lnteachers)) ///
> readlevel15 ///
> ///
> (ologit, omit(lnage povertysd pocsd lnstudents readall13 readall14 readall15 mathall
> 13 mathall14 mathall15 lnteachers)) ///
> mathlevel15 ///
> ///
> ///
> = numpdfs numwords numpages pocschoolcount ///
> inquiry_full_prop traditional_prop progressive_prop ///
> closerate publicdensity charterdensity ///
> i.urban statenum cmonum geodistrict, ///
> add(100) rseed(43) dots augment
note: variable lnage contains no soft missing (.) values; imputing nothing

```

Conditional models:

```

lnstudents: pmm lnstudents povertyschool numpdfs numwords numpages
               pocschoolcount inquiry_full_prop traditional_prop
               progressive_prop closerate publicdensity charterdensity i.urban
               statenum cmonum geodistrict , knn(5)
povertysd: pmm povertysd lnage pocsd lnteachers povertyschool readall15
               readall14 readall13 numpdfs numwords numpages pocschoolcount
               inquiry_full_prop traditional_prop progressive_prop closerate
               publicdensity charterdensity i.urban statenum cmonum
               geodistrict , knn(5)
pocsd: pmm pocsd lnage povertysd povertyschool numpdfs numwords
               numpages pocschoolcount inquiry_full_prop traditional_prop
               progressive_prop closerate publicdensity charterdensity i.urban
               statenum cmonum geodistrict , knn(5)
lnteachers: pmm lnteachers povertyschool i.readlevel15 mathall15
               i.mathlevel15 i.readlevel14 mathall14 i.mathlevel14
               i.readlevel13 mathall13 i.mathlevel13 numpdfs numwords numpages
               pocschoolcount inquiry_full_prop traditional_prop
               progressive_prop closerate publicdensity charterdensity i.urban
               statenum cmonum geodistrict , knn(5)
povertyschool: pmm povertyschool lnage povertysd pocsd numpdfs numwords
               numpages pocschoolcount inquiry_full_prop traditional_prop
               progressive_prop closerate publicdensity charterdensity i.urban
               statenum cmonum geodistrict , knn(5)
readall15: pmm readall15 lnage povertysd pocsd povertyschool i.readlevel15
               mathall15 i.mathlevel15 readall14 i.readlevel14 mathall14
               i.mathlevel14 readall13 i.readlevel13 mathall13 i.mathlevel13
               numpdfs numwords numpages pocschoolcount inquiry_full_prop
               traditional_prop progressive_prop closerate publicdensity
               charterdensity i.urban statenum cmonum geodistrict , knn(5)
readlevel15: ologit readlevel15 povertyschool i.mathlevel15 i.readlevel14
               i.mathlevel14 i.readlevel13 i.mathlevel13 numpdfs numwords
               numpages pocschoolcount inquiry_full_prop traditional_prop
               progressive_prop closerate publicdensity charterdensity i.urban
               statenum cmonum geodistrict , augment
mathall15: pmm mathall15 povertyschool i.readlevel15 i.mathlevel15
               i.readlevel14 mathall14 i.mathlevel14 i.readlevel13 mathall13
               i.mathlevel13 numpdfs numwords numpages pocschoolcount
               inquiry_full_prop traditional_prop progressive_prop closerate

```

```

publicdensity charterdensity i.urban statenum cmonum
geodistrict , knn(5)
mathlevel15: ologit mathlevel15 povertyschool i.readlevel15 i.readlevel14
i.mathlevel14 i.readlevel13 i.mathlevel13 numpdfs numwords
numpages pocschoolcount inquiry_full_prop traditional_prop
progressive_prop closerate publicdensity charterdensity i.urban
statenum cmonum geodistrict , augment
readall14: pmm readall14 lnage povertysd pocsd povertyschool readall15
i.readlevel15 mathall15 i.mathlevel15 i.readlevel14 mathall14
i.mathlevel14 readall13 i.readlevel13 mathall13 i.mathlevel13
numpdfs numwords numpages pocschoolcount inquiry_full_prop
traditional_prop progressive_prop closerate publicdensity
charterdensity i.urban statenum cmonum geodistrict , knn(5)
readlevel14: ologit readlevel14 povertyschool i.readlevel15 i.mathlevel15
i.mathlevel14 i.readlevel13 i.mathlevel13 numpdfs numwords
numpages pocschoolcount inquiry_full_prop traditional_prop
progressive_prop closerate publicdensity charterdensity i.urban
statenum cmonum geodistrict , augment
mathall14: pmm mathall14 povertyschool i.readlevel15 mathall15
i.mathlevel15 i.readlevel14 i.mathlevel14 i.readlevel13
mathall13 i.mathlevel13 numpdfs numwords numpages
pocschoolcount inquiry_full_prop traditional_prop
progressive_prop closerate publicdensity charterdensity i.urban
statenum cmonum geodistrict , knn(5)
mathlevel14: ologit mathlevel14 povertyschool i.readlevel15 i.mathlevel15
i.readlevel14 i.readlevel13 i.mathlevel13 numpdfs numwords
numpages pocschoolcount inquiry_full_prop traditional_prop
progressive_prop closerate publicdensity charterdensity i.urban
statenum cmonum geodistrict , augment
readall13: pmm readall13 lnage povertysd pocsd povertyschool readall15
i.readlevel15 mathall15 i.mathlevel15 readall14 i.readlevel14
mathall14 i.mathlevel14 i.readlevel13 mathall13 i.mathlevel13
numpdfs numwords numpages pocschoolcount inquiry_full_prop
traditional_prop progressive_prop closerate publicdensity
charterdensity i.urban statenum cmonum geodistrict , knn(5)
readlevel13: ologit readlevel13 povertyschool i.readlevel15 i.mathlevel15
i.readlevel14 i.mathlevel14 i.mathlevel13 numpdfs numwords
numpages pocschoolcount inquiry_full_prop traditional_prop
progressive_prop closerate publicdensity charterdensity i.urban
statenum cmonum geodistrict , augment
mathall13: pmm mathall13 povertyschool i.readlevel15 mathall15
i.mathlevel15 i.readlevel14 mathall14 i.mathlevel14
i.readlevel13 i.mathlevel13 numpdfs numwords numpages
pocschoolcount inquiry_full_prop traditional_prop
progressive_prop closerate publicdensity charterdensity i.urban
statenum cmonum geodistrict , knn(5)
mathlevel13: ologit mathlevel13 povertyschool i.readlevel15 i.mathlevel15
i.readlevel14 i.mathlevel14 i.readlevel13 numpdfs numwords
numpages pocschoolcount inquiry_full_prop traditional_prop
progressive_prop closerate publicdensity charterdensity i.urban
statenum cmonum geodistrict , augment

```

Performing chained iterations:

```

imputing m=1 through m=100 .....10.....20.....30.....40.....50..
> .....60.....70.....80.....90.....100 done

```

Multivariate imputation
Chained equations
Imputed: m=1 through m=100

```

Imputations =    100
added        =    100
updated      =     0

```

Initialization: **monotone**

```

Iterations =    1000
burn-in    =     10

```

```

povertyschool: predictive mean matching
povertysd: predictive mean matching
pocsd: predictive mean matching
lnage: predictive mean matching
readall13: predictive mean matching
mathall13: predictive mean matching
readall14: predictive mean matching
mathall14: predictive mean matching
readall15: predictive mean matching
mathall15: predictive mean matching
lnstudents: predictive mean matching
lnteachers: predictive mean matching
readlevel13: ordered logistic regression
mathlevel13: ordered logistic regression
readlevel14: augmented ordered logistic regression
mathlevel14: ordered logistic regression
readlevel15: ordered logistic regression
mathlevel15: augmented ordered logistic regression

```

Variable	Observations per <i>m</i>			
	Complete	Incomplete	Imputed	Total
povertyschool	5348	448	448	5796
povertysd	5768	28	28	5796
pocsd	5768	28	28	5796
lnage	5796	0	0	5796
readall13	4037	1759	1759	5796
mathall13	3830	1966	1966	5796
readall14	4960	836	836	5796
mathall14	4948	848	848	5796
readall15	5158	638	638	5796
mathall15	5142	654	654	5796
lnstudents	5784	1	1	5785
lnteachers	5595	152	152	5747
readlevel13	4037	1759	1759	5796
mathlevel13	3830	1966	1966	5796
readlevel14	4960	836	836	5796
mathlevel14	4948	848	848	5796
readlevel15	5158	638	638	5796
mathlevel15	5142	654	654	5796

(complete + incomplete = total; imputed is the minimum across *m* of the number of filled-in observations.)

Note: Right-hand-side variables (or weights) have missing values; model parameters estimated using listwise deletion.

Warning: the sets of predictors of the imputation model vary across imputations or iterations

```

41.
42.
43. ** -----
44. ** FINALIZE & SAVE DATA
45. ** -----
46.
47. * Recover variables using imputed natural-logarithmic versions:
48. replace age = exp(lnage)
   variable age was byte now float
   (38,064 real changes made)

```



```

max | .624036 .99 .99 .99 .99 .99 .99
> 1 sd | .1197104 .2347915 .2478774 .231485 .2437192 .2368318 .2469299 .324
> 4263

```

stats	povert~p	pocsd	povert~d	pctpdfs
N	5348	5768	5768	5796
mean	.5588811	.3348288	.1472116	.0054762
p50	.6018175	.31135	.1409	0
min	0	0	0	0
max	1	.992	.4495	1
sd	.3230693	.2004684	.075277	.0473971

m=1 data:

```

-> tabstat inquiry_full_log readall15 mathall15 readall14 mathall14 readall13 mathall1
> 3 pocschoolprop povertyschoolprop pocsd povertysd pctpdfs, stat(n mean median min ma
> x sd)

```

stats	in~l_log	reada~15	matha~15	reada~14	matha~14	reada~13	matha~13	pocs
ch~p								
N	5796	5784	5784	5784	5784	5786	5786	
> 5784								
mean	.1326029	.4731587	.4014315	.4826037	.4022614	.6028604	.5368545	.650
> 3792								
p50	.1294601	.47	.37	.47	.37	.65	.55	.738
> 5253								
min	0	.02	.02	.02	.02	.02	.02	
> 0								
max	.624036	.99	.99	.99	.99	.99	.99	
> 1								
sd	.1197104	.2349245	.2495253	.2314883	.2446586	.2354162	.247266	.324
> 4263								

stats	povert~p	pocsd	povert~d	pctpdfs
N	5784	5796	5796	5784
mean	.5551255	.3349206	.1472726	.0053739
p50	.5966306	.31135	.1409	0
min	0	0	0	0
max	1	.992	.4495	1
sd	.3249278	.2004833	.0753548	.0466637

72.

73. * Make sure original and imputed datasets have same summary stats for key variables:

```

74. mi xeq 0 1: tabstat inquiry_seed_log inquiry_narrow_log inquiry_full_log inquiry_ful
> l_nohands_log readall15 mathall15 pocschoolprop povertyschoolprop pocsd povertysd pc
> tpdfs, stat(n mean median min max sd)

```

m=0 data:

```

-> tabstat inquiry_seed_log inquiry_narrow_log inquiry_full_log inquiry_full_nohands_l
> og readall15 mathall15 pocschoolprop povertyschoolprop pocsd povertysd pctpdfs, stat
> (n mean median min max sd)

```

stats	in~d_log	in~w_log	in~l_log	in~s_log	reada~15	matha~15	pocsch~p	pove
> rt~p								
N	5796	5796	5796	5796	5158	5142	5784	
> 5348 mean	.0202511	.0672377	.1326029	.1167381	.4730109	.3976099	.6503792	.558
> 8811 p50	0	0	.1294601	.107191	.47	.36	.7385253	.601
> 8175 min	0	0	0	0	.02	.02	0	
> 0 max	.4908485	.4938436	.624036	.5206454	.99	.99	1	
> 1 sd	.0487966	.0876619	.1197104	.1135068	.2347915	.2478774	.3244263	.323
> 0693								

stats	pocsd	povert~d	pctpdfs
N	5768	5768	5796
mean	.3348288	.1472116	.0054762
p50	.31135	.1409	0
min	0	0	0
max	.992	.4495	1
sd	.2004684	.075277	.0473971

m=1 data:

```
> tabstat inquiry_seed_log inquiry_narrow_log inquiry_full_log inquiry_full_nohands_1
> og readall15 mathall15 pocschoolprop povertyschoolprop pocsd povertysd pctpdfs, stat
> (n mean median min max sd)
```

stats	in~d_log	in~w_log	in~l_log	in~s_log	reada~15	matha~15	pocsch~p	pove
> rt~p								
N	5796	5796	5796	5796	5784	5784	5784	
> 5784 mean	.0202511	.0672377	.1326029	.1167381	.4731587	.4014315	.6503792	.555
> 1255 p50	0	0	.1294601	.107191	.47	.37	.7385253	.596
> 6306 min	0	0	0	0	.02	.02	0	
> 0 max	.4908485	.4938436	.624036	.5206454	.99	.99	1	
> 1 sd	.0487966	.0876619	.1197104	.1135068	.2349245	.2495253	.3244263	.324
> 9278								

stats	pocsd	povert~d	pctpdfs
N	5796	5796	5784
mean	.3349206	.1472726	.0053739
p50	.31135	.1409	0
min	0	0	0
max	.992	.4495	1
sd	.2004833	.0753548	.0466637


```
75.
76. * Label variables:
77. labvars pocschoolprop "% students of color (school)" ///
  > povertyschoolprop "% students in poverty (school)" povertyschoolcount "# students in
  > poverty" ///
  > upperlevel "middle/high school" teacherratio "student/teacher ratio" ///
  > , alternate

78.
79. * Save data to disk
80. save "data/charter_schools_data_100_imputations.dta", replace
  (note: file data/charter_schools_data_100_imputations.dta not found)
  file data/charter_schools_data_100_imputations.dta saved

81. export delimited using "data/charter_schools_data_100_imputations.csv", replace
  (note: file data/charter_schools_data_100_imputations.csv not found)
  file data/charter_schools_data_100_imputations.csv saved

82.
83. log close
    name: <unnamed>
    log: /hdir/0/jhaber/Projects/charter_data/sorting-schools-2019/logs/multiple_i
> mputation_100_101019.smcl
    log type: smcl
    closed on: 11 Oct 2019, 08:31:15
```
