

```
*****;
** KRUEGER MI PRESENTATION *****;
*****;
. tab rac3, gen(race);
```

rac3	Freq.	Percent	Cum.
1	34,369	83.61	83.61
2	5,604	13.63	97.25
3	1,131	2.75	100.00
Total	41,104	100.00	

```
. rename race1 white;
. rename race2 black;
. rename race3 other;

. gen ag2=ag*ag;

. gen sleep2=sleep*sleep;
(212 missing values generated)

. gen bmi2=bmi*bmi;
(495 missing values generated)

. summ bmi faminc white black other male ag ag2 inc sleep sleep2, sep(0);
```

Variable	Obs	Mean	Std. Dev.	Min	Max
bmi	40609	25.05989	4.81942	13.12898	98.6046
faminc	35169	30.86098	21.49091	.5	68.645
white	41104	.8361473	.3701463	0	1
black	41104	.1363371	.3431503	0	1
other	41104	.0275156	.1635821	0	1
male	41104	.4177696	.4931978	0	1
ag	41104	45.28861	18.29368	18	99
ag2	41104	2385.709	1852.785	324	9801
inc	40354	.5895326	.4919247	0	1
sleep	40892	7.288907	1.348323	0	23
sleep2	40892	54.9461	21.35609	0	529

```
. corr bmi faminc white black other male ag inc sleep ;
(obs=34742)
```

	bmi	faminc	white	black	other	male	ag	inc	sleep
bmi	1.0000								
faminc	-0.0618	1.0000							
white	-0.0911	0.1728	1.0000						
black	0.1261	-0.1845	-0.8916	1.0000					
other	-0.0553	-0.0059	-0.3925	-0.0666	1.0000				
male	0.1010	0.1097	0.0482	-0.0625	0.0206	1.0000			
ag	0.1082	-0.1055	0.0519	-0.0290	-0.0555	-0.0392	1.0000		
inc	-0.0511	0.7477	0.1723	-0.1835	-0.0070	0.1137	-0.1416	1.0000	
sleep	-0.0359	-0.0644	-0.0288	0.0385	-0.0147	-0.0229	0.0348	-0.0685	1.0000

```
. reg bmi faminc black other male ag ag2;
```

Source	SS	df	MS	Number of obs =	34863
Model	62489.1639	6	10414.8607	F(6, 34856) =	481.62
Residual	753751.432	34856	21.6247255	Prob > F =	0.0000
				R-squared =	0.0766
				Adj R-squared =	0.0764
Total	816240.596	34862	23.4134759	Root MSE =	4.6502

bmi	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
faminc	-.0224783	.0012544	-17.92	0.000	-.024937 - .0200196
black	1.638977	.0754914	21.71	0.000	1.491012 1.786943
other	-1.283572	.1503717	-8.54	0.000	-1.578305 -.9888386
male	1.130308	.0507729	22.26	0.000	1.030791 1.229824
ag	.3113446	.0082619	37.68	0.000	.295151 .3275383
ag2	-.0028666	.0000829	-34.59	0.000	-.003029 -.0027041
_cons	17.86139	.1795452	99.48	0.000	17.50948 18.21331

```
. ice
> bmi bmi2 faminc black other male ag ag2 inc sleep sleep2 m.exoth
> , seed(111511) cycles(100) m(5) saving($work/imputed1.dta, replace) trace($work/tracel.dta)
genmiss(miss)
> cmd(faminc:nbreg, sleep:reg)
> passive(
> sleep2: sleep*sleep
> \      bmi2: bmi*bmi
> )
> eqdrop(
> inc: faminc)
> ;
```

```
=> xi: ice bmi bmi2 faminc black other male ag ag2 inc sleep sleep2 exoth i.exoth, cmd(exoth:mlogit,
faminc:nbreg, sleep:reg) substitute(exoth:i.ex
> oth) seed(111511) cycles(100) m(5) saving(C:\DATA\Analyses\GradStat/imputed1.dta, replace)
trace(C:\DATA\Analyses\GradStat/tracel.dta) genm
> iss(miss) passive( sleep2: sleep*sleep \      bmi2: bmi*bmi ) eqdrop( inc: faminc)
```

```
i.exoth      _Iexoth_1-3      (naturally coded; _Iexoth_1 omitted)
```

#missing values	Freq.	Percent	Cum.
0	34,216	83.24	83.24
1	5,370	13.06	96.31
2	1,147	2.79	99.10
3	270	0.66	99.75
4	75	0.18	99.94
5	18	0.04	99.98
6	7	0.02	100.00
7	1	0.00	100.00
Total	41,104	100.00	

Variable	Command	Prediction equation
bmi	regress	faminc black other male ag ag2 inc sleep sleep2 _Iexoth_2
bmi2		_Iexoth_3
		[Passively imputed from bmi*bmi]
faminc	nbreg	bmi bmi2 black other male ag ag2 inc sleep sleep2
		_Iexoth_2 _Iexoth_3
black		[No missing data in estimation sample]
other		[No missing data in estimation sample]
male		[No missing data in estimation sample]
ag		[No missing data in estimation sample]
ag2		[No missing data in estimation sample]

```

    inc | logit | bmi bmi2 black other male ag ag2 sleep sleep2
        |     | _Iexoth_2 _Iexoth_3
sleep | reg   | bmi bmi2 faminc black other male ag ag2 inc _Iexoth_2
        |     | _Iexoth_3
sleep2 |      | [Passively imputed from sleep*sleep]
exoth  | mlogit | bmi bmi2 faminc black other male ag ag2 inc sleep sleep2
_Iexoth_2 |      | [Passively imputed from (exoth==2)]
_Iexoth_3 |      | [Passively imputed from (exoth==3)]

```

Imputing

```

[Perfect prediction detected: using augmlogit to impute exoth]
.....1
[Perfect prediction detected: using augmlogit to impute exoth]
.....2
[Perfect prediction detected: using augmlogit to impute exoth]
.....3
[Perfect prediction detected: using augmlogit to impute exoth]
.....4
[Perfect prediction detected: using augmlogit to impute exoth]
.....5
file C:\DATA\Analyses\GradStat\imputed1.dta saved

```

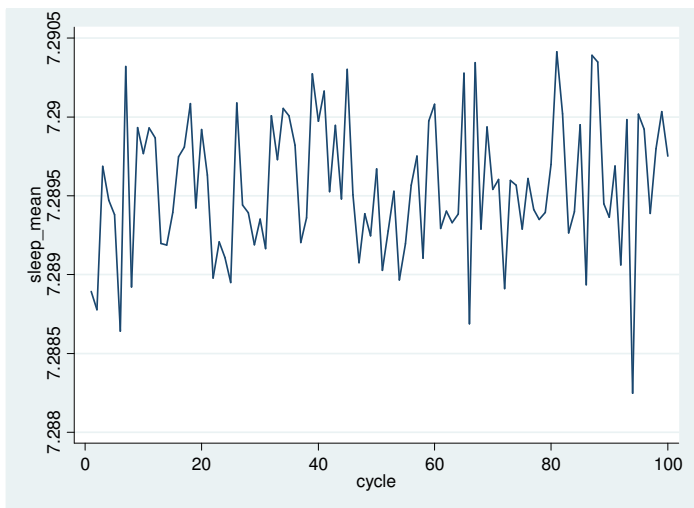
```

. use $work/tracel.dta, clear

. sort cycle

. line sleep_mean cycle;

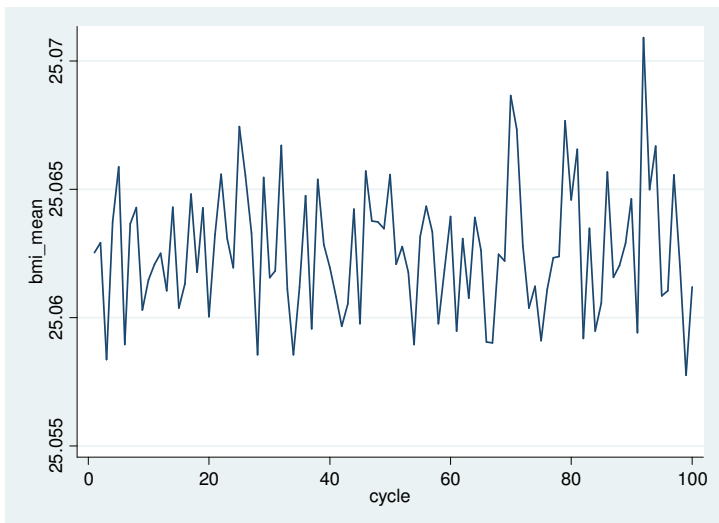
```



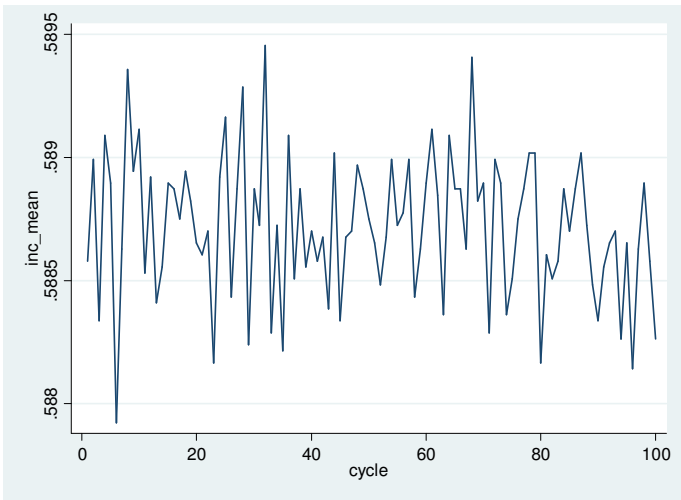
```

. line bmi_mean cycle;

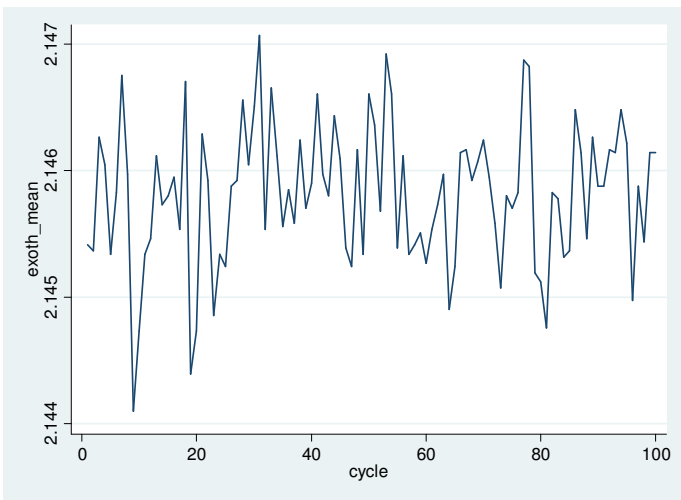
```



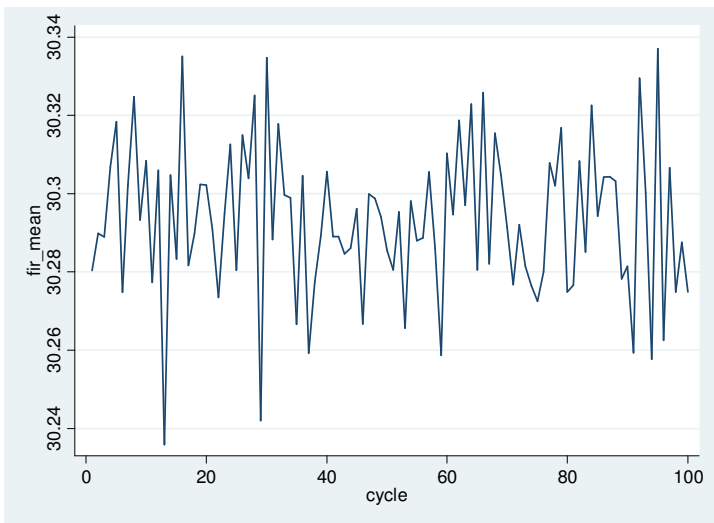
```
. line inc_mean cycle;
```



```
. line exoth_mean cycle;
```



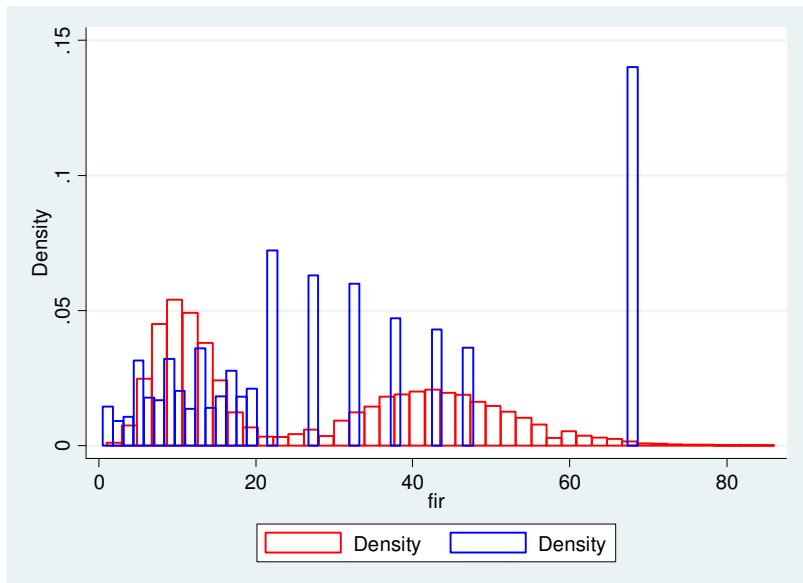
```
. line faminc_mean cycle;
```



```
. use $work/imputed1.dta, clear
```

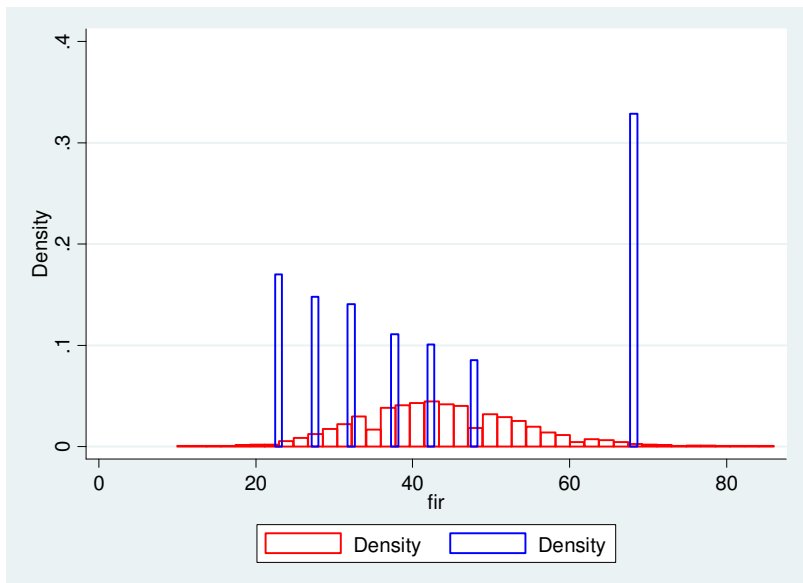
```
** RED IS IMPUTED BLUE IS OBSERVED **;
```

```
. twoway histogram faminc if missfaminc==1 & _mj>0, fcolor(none) lcolor(red) || histogram faminc if  
missfaminc==0 & _mj>0, fcolor(none) lcolor(blue);
```



```
** FOR INCOME ABOVE 20k, RED IS IMPUTED **;
```

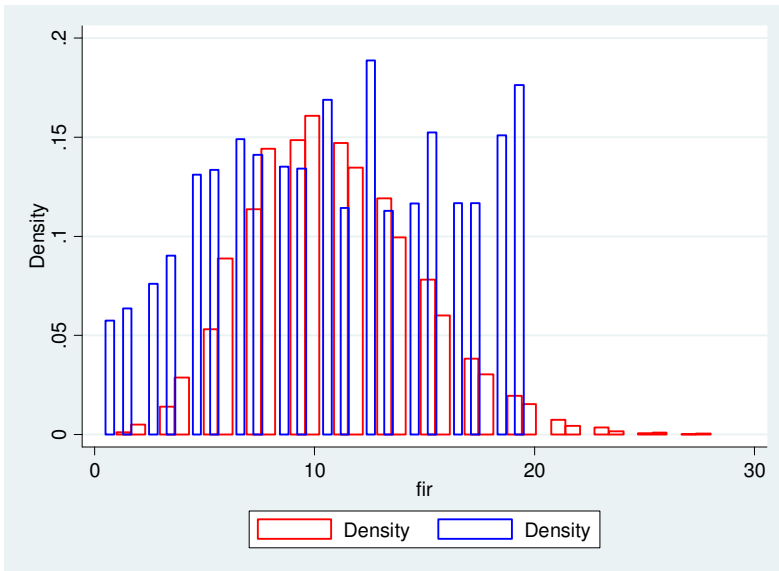
```
. twoway histogram faminc if missfaminc==1 & _mj>0 & inc==1, fcolor(none) lcolor(red) || histogram faminc  
if missfaminc==0 & _mj>0 & inc==1, fcolor(none)  
> lcolor(blue);
```



```

** FOR INCOME BELOW 20k, RED IS IMPUTED **;
. twoway histogram faminc if missfaminc==1 & _mj>0 & inc==0, fcolor(none) lcolor(red) || histogram faminc
if missfaminc==0 & _mj>0 & inc==0, fcolor(none)
> lcolor(blue);

```



```

. drop miss*;

. mi import ice, automatic clear
(7452 m=0 obs. now marked as incomplete)

. mi query;
data mi set flong, M = 5
last mi update 12nov2011 19:41:40, approximately 4 minutes ago

. mi describe;

Style:  flong
      last mi update 12nov2011 19:41:40, approximately 4 minutes ago

Obs.:  complete      33,652
      incomplete     7,452  (M = 5 imputations)
      -----
      total          41,104

Vars.:  imputed:  18; mss(35) ed(55) srh(105) inc(750) faminc(5935) high(156) heavy(416) sleep(212)
stressa(862) stressc(863) exreg(375)
      exoth(814) smoke(438) bmi(495) sleep2(212) bmi2(495) _Iexoth_2(814) _Iexoth_3(814)

passive:  0

regular:  0

system:   3; _mi_m _mi_id _mi_miss

(there are 26 unregistered variables)

```

```
. mi estimate: reg bmi faminc black other male ag ag2;
```

```
Multiple-imputation estimates      Imputations      =          5
Linear regression                 Number of obs    =       41104
                                   Average RVI      =        0.0172
                                   Complete DF     =       41097
DF adjustment:  Small sample      DF:      min    =       3133.17
                                   avg          =      12881.55
                                   max          =      39533.12
Model F test:      Equal FMI      F(   6,22542.3) =       557.22
Within VCE type:  OLS            Prob > F       =        0.0000
```

bmi	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
faminc	-.0221756	.0011864	-18.69	0.000	-.0245011 - .0198501
black	1.647164	.0686351	24.00	0.000	1.512634 1.781694
other	-1.244473	.1427492	-8.72	0.000	-1.524361 -.9645856
male	1.137109	.0467463	24.33	0.000	1.045485 1.228733
ag	.3067233	.0075736	40.50	0.000	.2918751 .3215716
ag2	-.0028182	.0000756	-37.26	0.000	-.0029665 -.0026699
_cons	17.90082	.1654968	108.16	0.000	17.5764 18.22525

```
. mi test ag ag2;
note: assuming equal fractions of missing information
```

- (1) ag = 0
- (2) ag2 = 0

```
F( 2,2944.2) = 935.40
Prob > F = 0.0000
```

```
. log close; exit;
```

```
***** RE-POST COEFFICIENT TABLES *****
```

```
COEFFICIENT TABLE FROM IMPUTED DATA (N=41,104; M=5)
```

bmi	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
faminc	-.0221756	.0011864	-18.69	0.000	-.0245011 - .0198501
black	1.647164	.0686351	24.00	0.000	1.512634 1.781694
other	-1.244473	.1427492	-8.72	0.000	-1.524361 -.9645856
male	1.137109	.0467463	24.33	0.000	1.045485 1.228733
ag	.3067233	.0075736	40.50	0.000	.2918751 .3215716
ag2	-.0028182	.0000756	-37.26	0.000	-.0029665 -.0026699
_cons	17.90082	.1654968	108.16	0.000	17.5764 18.22525

```
COEFFICIENT TABLE FROM NON-IMPUTED DATA (N=34,863)
```

bmi	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
faminc	-.0224783	.0012544	-17.92	0.000	-.024937 - .0200196
black	1.638977	.0754914	21.71	0.000	1.491012 1.786943
other	-1.283572	.1503717	-8.54	0.000	-1.578305 -.9888386
male	1.130308	.0507729	22.26	0.000	1.030791 1.229824
ag	.3113446	.0082619	37.68	0.000	.295151 .3275383
ag2	-.0028666	.0000829	-34.59	0.000	-.003029 - .0027041
_cons	17.86139	.1795452	99.48	0.000	17.50948 18.21331

```
NOTES:
```

1. General conclusions are the same—coefficients are similar, all items are significant
2. Imputed data have smaller standard errors and larger t-statistics due to larger sample size, even for variables that were not imputed
 - a. Some t-statistics are almost 3 units greater for imputed data—a substantial difference if we were working with a smaller data set