```
 1 . use "E:\16GBBACKUPUSB\BACKUP_USB_SEPTEMBER2014\May Baydoun_folder\UK_BIOBANK_PROJECT\UKB_PAPER8D_LE8PROTDEM\DAT

 2 .
 3 .
 4 . ********************REVISION*****************************
 5 .
 6 . capture drop samplefinal_initial

 7 . gen samplefinal_initial=.
(502,409 missing values generated)

 8 . replace samplefinal_initial=1 if sample_final==1
(28,974 real changes made)

 9 . replace samplefinal_initial=0 if samplefinal_initial==. & AGE>=50
(355,668 real changes made)

10 .
11 . save, replace
file E:\16GBBACKUPUSB\BACKUP_USB_SEPTEMBER2014\May Baydoun_folder\UK_BIOBANK_PROJECT\UKB_PAPER8D_LE8PROTDEM\DATA\

12 .
13 . logistic samplefinal_initial AGE SEX NonWhite SES householdsize

Logistic regression                                     Number of obs = 356,090
                                                        LR chi2(5)    =  116.02
                                                        Prob > chi2   =  0.0000
Log likelihood = -100393.41                             Pseudo R2     =  0.0006

------------------------------------------------------------------------------
samplefinal_initial | Odds ratio   Std. err.      z    P>|z|     [95% conf. interval]
--------------------+---------------------------------------------------------
                AGE |    1.00927   .001188     7.84   0.000     1.006945    1.011601
                SEX |   .9607232   .0118483   -3.25   0.001     .9377793    .9842284
           NonWhite |   1.053242   .0319645    1.71   0.087     .9924192    1.117792
                SES |   .9550169   .0087021   -5.05   0.000     .9381125    .9722259
      householdsize |   1.008635   .0053493    1.62   0.105     .9982045    1.019174
              _cons |   .0525926   .0041493  -37.33   0.000     .0450576    .0613875
------------------------------------------------------------------------------
Note: _cons estimates baseline odds.

14 .
15 . tab samplefinal_initial SEX, row col chi

+----------------------+
| Key                  |
|----------------------|
|     frequency        |
|  row percentage      |
| column percentage    |
+----------------------+
```

samplefinal_initial	SEX 1	2	Total
0	**162,154**	**193,494**	**355,648**
	45.59	**54.41**	**100.00**
	92.18	**92.71**	**92.47**
1	**13,759**	**15,215**	**28,974**
	47.49	**52.51**	**100.00**
	7.82	**7.29**	**7.53**
Total	**175,913**	**208,709**	**384,622**
	45.74	**54.26**	**100.00**
	100.00	**100.00**	**100.00**

Pearson chi2(**1**) = **38.7016** Pr = **0.000**

16 . tab samplefinal_initial NonWhite, row col chi

Key
frequency *row percentage* *column percentage*

samplefinal_initial	ethnic grouping 3 white	Non-White	Total
0	**339,112**	**16,556**	**355,668**
	95.35	**4.65**	**100.00**
	92.45	**92.85**	**92.47**
1	**27,699**	**1,275**	**28,974**
	95.60	**4.40**	**100.00**
	7.55	**7.15**	**7.53**
Total	**366,811**	**17,831**	**384,642**
	95.36	**4.64**	**100.00**
	100.00	**100.00**	**100.00**

Pearson chi2(**1**) = **3.9223** Pr = **0.048**

17 . reg samplefinal_initial AGE

Source	SS	df	MS		
Model	**.799069719**	**1**	**.799069719**	Number of obs	= **384,619**
Residual	**26790.5405**	**384,617**	**.069655113**	F(1, 384617)	= **11.47**
				Prob > F	= **0.0007**
				R-squared	= **0.0000**
				Adj R-squared	= **0.0000**
Total	**26791.3396**	**384,618**	**.069657009**	Root MSE	= **.26392**

samplefinal~l	Coefficient	Std. err.	t	P>|t|	[95% conf.	interval]
AGE	**.0002645**	**.0000781**	**3.39**	**0.001**	**.0001115**	**.0004176**
_cons	**.0593078**	**.0047501**	**12.49**	**0.000**	**.0499977**	**.0686178**

```
18 . reg samplefinal_initial SES
```

Source	SS	df	MS	Number of obs	=	357,646
				F(1, 357644)	=	31.54
Model	2.34773895	1	2.34773895	Prob > F	=	0.0000
Residual	26624.3796	357,644	.074443803	R-squared	=	0.0001
				Adj R-squared	=	0.0001
Total	26626.7273	357,645	.07445016	Root MSE	=	.27284

samplefina~l	Coefficient	Std. err.	t	P>|t|	[95% conf.	interval]
SES	-.003693	.0006576	-5.62	0.000	-.004982	-.0024041
_cons	.0808345	.0004573	176.75	0.000	.0799381	.0817309

```
19 . reg samplefinal_initial householdsize
```

Source	SS	df	MS	Number of obs	=	381,025
				F(1, 381023)	=	0.92
Model	.064350524	1	.064350524	Prob > F	=	0.3386
Residual	26770.6874	381,023	.07026003	R-squared	=	0.0000
				Adj R-squared	=	-0.0000
Total	26770.7517	381,024	.070260014	Root MSE	=	.26507

samplefinal~l	Coefficient	Std. err.	t	P>|t|	[95% conf.	interval]
householdsize	-.0003415	.0003568	-0.96	0.339	-.0010409	.0003579
_cons	.0768037	.0009041	84.95	0.000	.0750316	.0785758

```
20 .
21 . bysort samplefinal_initial: su AGE SES householdsize
```

-> samplefinal_initial = 0

Variable	Obs	Mean	Std. dev.	Min	Max
AGE	355,645	60.56711	5.444151	50.00137	73.69199
SES	328,672	-.0464156	.6934405	-2.704226	1.807114
households~e	352,051	2.230273	1.205543	1	60

-> samplefinal_initial = 1

Variable	Obs	Mean	Std. dev.	Min	Max
AGE	28,974	60.67986	5.50548	50.00137	70.98973
SES	28,974	-.0702908	.6971128	-2.528219	1.757327
households~e	28,974	2.223235	1.177014	1	61

-> samplefinal_initial = .

Variable	Obs	Mean	Std. dev.	Min	Max
AGE	117,767	45.46717	2.742084	37.41821	49.99863
SES	114,220	.0681585	.7219972	-2.539996	1.749102
households~e	116,815	3.099396	1.479142	1	100

```
22 .
23 . save, replace
   file E:\16GBBACKUPUSB\BACKUP_USB_SEPTEMBER2014\May Baydoun_folder\UK_BIOBANK_PROJECT\UKB_PAPER8D_LE8PROTDEM\DATA\

24 .
25 . capture log close
```