```
                                        __  ___  ___  __  ___(R)
                                       /__  /   /__/ /   /__/
                                      ___/ /   /__/ /   /__/
                                         Statistics/Data analysis
```

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

17 .
   end of do-file

18 .