

log using "D:\Jason\workshop\Introduction to Stata Analysis\Introduction to Stata Analysis2.log", replace

```
*****
* This command file demonstrate the steps of determining if gender is a significant predictor of health.
* Respondent's age is used as another independent variable and the control variable is whether respondents
* live in a city or rural area
*****
```

use <http://www.stata-press.com/data/r13/nhanes2f>, clear

```
*****
* check the data and the number of observations
*****
```

des, short
count

```
*****
* Select variables
*****
```

keep health agegrp female rural psuid finalwgt stratid

```
*****
* Check the distribution of variables
*****
```

des health agegrp female rural psuid finalwgt stratid
sum health agegrp female rural psuid finalwgt stratid
tab1 health agegrp female rural, mis

```
*****
* Create missing values on two variable (i.e., health and agegrp)
* for demonstrating how to use the -mi impute command
*****
```

set seed 98034
generate u1 = runiform()
tab1 u1, mis
replace health =. if u1 >= .9487244
replace agegrp =. if u1 >= .98

```
*****
* Step #1: multiple imputation
*****
```

mi set flong
mi misstable summarize health agegrp female rural psuid finalwgt stratid

mi register imputed health agegrp
mi register regular female rural
mi impute chained (ologit) health (ologit) agegrp = i.female i.rural, add(5) rseed(98034)

* mi passive: generate age_female = agegrp*female
* label variable age_female "interaction term of agegroup and female"

```
*****
* Step #2: Set the -svy-command
*****
```

mi svyset psuid [pweight=finalwgt], strata(stratid)

```

*****
* Step #3: Obtain descriptive statistics
*****
mi estimate: svy: mean health agegrp
mi estimate: svy: proportion female rural

*****
* Step 4: Perform a series of regression analyses to put the theoretical questions to the test.
* The research question is if being female is associated with an increased level of self-reported health
* after two variables (agegrp and rural) are in the model
* The post-estimation test also test if both respondent's age and gender predict the outcome variable
* and if age and gender have the same association with the outcome variable
*****
mi estimate: svy: ologit health agegrp
mi estimate: svy: ologit health agegrp female
mi estimate, saving(d:\temp\full.dta, replace): svy: ologit health agegrp female rural

*****
* post-estimation test
*****
mi test agegrp female
mi estimate (diff: _b[agegrp]-_b[female]) using d:\temp\full.dta

*****
* The results are reasonable, so no Steps 5 and 6
*****

*****
* Step 7: Sensitivity test by using a new operationalization of health in analysis
*****

clonevar health2 = health
recode health2 (1/2=1) (3=2) (4/5=3)
tab2 health health2
label variable health2 "recoded Health into a three-category variable"

*****
* reanalyze the data, using the recoded Health2 variable as the outcome variable
*****

mi estimate: svy: ologit health2 agegrp
mi estimate: svy: ologit health2 agegrp female
mi estimate, saving(d:\temp\full2.dta, replace): svy: ologit health2 agegrp female rural

*****
* post-estimation test
*****
mi test agegrp female

mi estimate (diff: _b[agegrp]-_b[female]) using d:\temp\full2.dta

log close

```