Applied Statistical Analysis with Missing Data
Friday morning – a light exercise

The data set `friday1.dta` is an artificial data set containing three variables `y` (the response), `x` (a continuous explanatory variable) and `sex` (1= man, 0=woman).
The focus is on association between `y` and `x`, but a large proportion of `x` is missing.

Convince you self that the data **could be** MCAR by running the commands:
```
generate xmiss=missing(x)
prtest xmiss, by(sex)
ttest y      , by(xmiss)
binreg xmiss y b0.sex, or
```

Discuss what you might gain from using multiple imputation methods on this data set.

We want to impute the missing `x`'s by the model `regress y x.`
Check the validity of the models by the usual qnorm-plot of residuals and a plot of the residuals against `y`.

Impute the missing x values (5 times) by the code:
```
mi set flong
mi register imputed x
mi register regular y sex
mi describe
mi impute chained (regress) x= y , add(5) rseed(234567)
```

Part A   the `y-x` association

Estimate `y-x` association based on complete data:
```
regress y x if _mi_m==0
```
Note the number of data included in the analysis and fill in the estimates and confidence intervals in the table on the next side.

Estimate `y-x` association based on the imputed data:
```
mi estimate: regress y x
```
Note the number of data included in the analysis and fill in the estimates and confidence intervals in the table.
Compare these with the complete data analysis above. What have you gained?

Part B   the `y-x` association adjusted for `sex` and the `y-sex` association adjusted for `x`

Estimate the two mutually adjusted associations based on the complete data:
```
regress y x b0.sex if _mi_m==0
```
Fill in the estimates and their confidence intervals in the table.

Repeat the analysis on the imputed data
```
mi estimate: regress y x b0.sex
```
and compare with the results based on the complete data.

Part C   the y-x modification by sex

Estimate the model with interaction between x and sex based on the complete data:

```
regress y x b0.sex b0.sex#c.x if _mi_m==0
```

Fill in the estimates and their confidence intervals in the table.

Repeat the analysis on the imputed data

```
mi estimate: regress y x b0.sex b0.sex#c.x
```

and compare with the results based on the complete data.

| | Complete case | | | imputed by (regress) x= y | | |
	estimate	low	high	estimate	low	high
Part A						
Slope x						
Constant						
Part B						
Slope x						
Men vs Women						
Constant						
Part C						
Slope x (Women)						
Men vs Women (x=0)						
Slope Men - slope Women						
Constant						