

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.

Part A the `y-x` association

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
```

Estimate `y-x` association based on complete data:

```
regress y x
```

Note the number of data included in the analysis, the estimated slope, its standard error and confidence interval.

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

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)
```

Estimate `y-x` association based on the imputed data:

```
mi estimate: regress y x
```

Note the number of data included in the analysis, the estimated slope, its standard error and confidence interval. 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 association based on the complete data:

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

Note the relevant estimates and their confidence intervals.

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 based on the complete data:

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

Note the interaction/effect modification and its confidence interval.

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.