Amelia

Hao Wang

January 22, 2017

```
setwd("D:/Dropbox/2017 Spring/ 603/lab/Lab 1")
#setwd('~/Dropbox/2017 Spring/ 603/lab/Lab 1')
getwd()
```

```
## [1] "D:/Dropbox/2017 Spring/ 603/lab/Lab 1"
```

Additional Training (if time permitted): Amelia

The following materials borrow heavily from UC Berkeley *R for Data Science* workshop. https://github.com/dlab-berkeley/R-for-Data-Science

```
large <-read.csv('https://raw.githubusercontent.com/haowang666/R-for-Data-Science/master/data/large.csv
summary(large)
```

```
##        a                    b                c          
##  Min.   :-33.98426   Min.   :-13.4   Min.   :-249998.64  
##  1st Qu.: -6.71903   1st Qu.:128.6   1st Qu.:-141005.65  
##  Median :  0.41681   Median :256.9   Median : -63498.56  
##  Mean   :  0.00176   Mean   :252.2   Mean   : -83954.09  
##  3rd Qu.:  7.00630   3rd Qu.:377.5   3rd Qu.: -15748.98  
##  Max.   : 35.33306   Max.   :513.3   Max.   :     11.77  
##  NA's   :45          NA's   :45      NA's   :45
```

```
nrow(na.omit(large))
```

```
## [1] 871
```

Package **Amelia** can impute missing when missingness is low and N is large

```
require(Amelia)
```

```
## Loading required package: Amelia
```

```
## Warning: package 'Amelia' was built under R version 3.3.2
```

```
## Loading required package: Rcpp
```

```
## ##
## ## Amelia II: Multiple Imputation
## ## (Version 1.7.4, built: 2015-12-05)
## ## Copyright (C) 2005-2017 James Honaker, Gary King and Matthew Blackwell
## ## Refer to http://gking.harvard.edu/amelia/ for more information
## ##
```

```
a <- amelia(large, m=1)
```

```
## -- Imputation 1 --
##
##   1  2  3
```

```
print(a)
```

```
##
## Amelia output with 1 imputed datasets.
## Return code:  1
## Message:  Normal EM convergence.
##
## Chain Lengths:
## --------------
## Imputation 1:  3
```

Amelia returns a list, when the first item is a list of your imputations. We only did one, so:

```
large.imputed <- a[[1]][[1]]
summary(large.imputed)
```

```
##        a                    b               c
##  Min.   :-33.98426   Min.   :-13.4   Min.   :-249999
##  1st Qu.: -6.83242   1st Qu.:126.1   1st Qu.:-140641
##  Median :  0.50371   Median :252.1   Median : -64017
##  Mean   :  0.06687   Mean   :250.4   Mean   : -83246
##  3rd Qu.:  7.21857   3rd Qu.:376.3   3rd Qu.: -15561
##  Max.   : 35.33306   Max.   :546.7   Max.   :  68074
```