

## Multilevel Multiple Imputation Example

Below I used the Keller and Enders (2023) program Blimp to illustrate the FCS multilevel multiple imputation method for the ECLS data (Enders, Keller, & Levy, 2017).<sup>1</sup> Blimp is used for the first step to impute 20 data sets and then R `lme4` is used to obtain the pooled multilevel estimate for the multiple imputed data sets. The model used three variables from the ECLS data set.<sup>2</sup> For just these three variables, roughly 20% of the cases would have been lost had listwise deletion been used. The model tests the relationship between the predictors, interpersonal problems and personal interpersonal skills, and the outcome, reading scores.

Statistics				
		reading X5 READING IRT SCALE SCORE-K2 DATA FILE	intprob X5 TEACHER REPORT INTERN PROB BEHAVIORS	perskills X5 TEACHER REPORT INTERPERS ONAL
N	Valid	4484	4118	3966
	Missing	247	613	765

### Blimp

The Blimp package can be used to generate multiple imputed data sets and save them. There are also a number of diagnostic features, which I do not illustrate here. Data must be in comma-separated format (.csv) with no labels. The ordinal line lists predictors and any additional auxiliary variables desired. Missing codes are -99 for this data set. There are 20 data sets generated with the `NIMPS` command. Burn-in is set to 1500 here, which indicates that 1500 MCMC iterations will be used prior to producing the first data set. The between-imputation interval (`THIN`) is 1500 iterations, and the `SEED` sets a start value for the random number generator to allow for reproducing the same results. Data sets can be output into the same file (`stacked`) or into multiple files (`separate`).

```
DATA: c:/jason/spsswin/mlrclass/ecls2011 missing.csv;
VARIABLES: childid schoolid reading lang intprob perskills bmi
           percmeal percmin schsize;
ORDINAL: perskills intprob reading;
MISSING: -99;
MODEL: schoolid ~ reading perskills intprob reading*perskills;
NIMPS: 20;
BURN: 1500;
THIN: 1500;
SEED: 90291;
OUTFILE: c:/jason/blimp/ecls2011 missing stacked.csv;
OPTIONS:
```

### R

In R, the stacked data set containing 20 imputed data sets sorted by imputation number (`imp`) is read in. The output file for the Blimp run gives the order of the variables and those must match the order specified on the names line. The `mitml` and the `lme4` libraries are needed. To keep it simple, I had only two predictors and did not include any random slopes. The `with` function asks `mitml` to pool the 20 data sets when running `lmer`. `coef` is used to obtain the pooled coefficients, `vcov` to obtain the variances and covariances of estimates, and `confint` to obtain the pooled confidence intervals. Likelihood ratio tests are available through the `testModel()` function.

<sup>1</sup> Download the program for free at <http://www.appliedmissingdata.com/multilevel-imputation.html>.

<sup>2</sup> It would be preferable to add other variables as auxiliary variables or include in the model under most circumstances.

```
> library(mitml)
>
> # read stacked blimp file
> impdata <- read.csv(file = "c:/jason/blimp/ec1s2011 missing stacked.csv", head = FALSE, sep = "
",)
> names(impdata) = c("imp", "childid", "schoolid", "reading", "lang", "intprob", "perskills", "bm
i",
+                   "percmeal", "percmin", "schsize")
>
> # split stacked data into separate files
> implist <- split(impdata, impdata$imp)
> implist <- as.mitml.list(implist)
>
> # multilevel regression with lm
> require(lme4)
> model <- with(implist, lmer(reading ~ intprob + perskills + (1|schoolid), REML = TRUE))
>
> # pooled estimates
> impmodel <- testEstimates(model, df.com = NULL)
> coef(impmodel)
(Intercept)      intprob      perskills
      83.71415      -4.19242       3.77882
> vcov(impmodel)
              (Intercept)      intprob      perskills
(Intercept)  3.0585199 -0.59188355 -0.63429272
intprob      -0.5918836  0.25231744  0.07054931
perskills    -0.6342927  0.07054931  0.16756929
> confint(impmodel)
              2.5 %      97.5 %
(Intercept) 80.284204 87.144105
intprob     -5.177086 -3.207755
perskills    2.975689  4.581952
```

#### References

Enders, C. K., Keller, B. T., & Levy, R. (2017). A chained equations imputation approach for multi-level data with categorical and continuous variables. *Psychological Methods*, Advance online publication. <http://dx.doi.org/10.1037/met0000148>.

Keller, B. T., & Enders, C. K. (2023). *Blimp user's guide (Version 3)*. Retrieved from [www.appliedmissingdata.com/blimp](http://www.appliedmissingdata.com/blimp)