Combining multiple imputations

Thomas Lumley

July 26, 2004

Carlin *et al.* (2003) illustrate the use of their Stata texttt for multiple imputations with data from a cohort study of adolescent health. Five sets of imputations were done, separately for male and female participants. The resulting datasets are in `mitools/dta`.

First we read all the datasets into R, using `read.dta` from the `foreign` package.

```
> library(mitools)
> data.dir <- system.file("dta", package = "mitools")
> library(foreign)
> women <- imputationList(lapply(list.files(data.dir, pattern = "f.\\.dta",
+     full = TRUE), read.dta, warn.missing.labels = FALSE))
> men <- imputationList(lapply(list.files(data.dir, pattern = "m.\\.dta",
+     full = TRUE), read.dta, warn.missing.labels = FALSE))
```

We now combine the imputations for men and women, first defining a `sex` variable

```
> women <- update(women, sex = 0)
> men <- update(men, sex = 1)
> all <- rbind(women, men)
> all

MI data with 5 datasets
Call: rbind(...)

> colnames(all)
```

```
 [1] "id"      "wave"    "mmetro"  "parsmk"  "drkfre"  "alcdos"
 [7] "alcdhi"  "smk"     "cistot"  "mdrkfre" "sex"
```

Now tabulate drinking frequency by sex

```
> with(all, table(sex, drkfre))

[[1]]
   drkfre
sex Non drinker not in last wk <3 days last wk >=3 days last wk
  0 282         201            105              12
  1 207         194            134              35

[[2]]
   drkfre
sex Non drinker not in last wk <3 days last wk >=3 days last wk
  0 282         195            109              14
  1 200         200            132              38

[[3]]
   drkfre
sex Non drinker not in last wk <3 days last wk >=3 days last wk
  0 278         202            109              11
  1 209         194            131              36

[[4]]
   drkfre
sex Non drinker not in last wk <3 days last wk >=3 days last wk
  0 284         188            114              14
  1 203         206            128              33

[[5]]
   drkfre
sex Non drinker not in last wk <3 days last wk >=3 days last wk
  0 288         191            109              12
  1 206         192            136              36

attr(,"call")
with.imputationList(all, table(sex, drkfre))
```

and define a new 'regular drinking' variables.

```
> all <- update(all, drkreg = as.numeric(drkfre) > 2)
> with(all, table(sex, drkreg))
```

```
[[1]]
   drkreg
sex FALSE TRUE
  0 483   117
  1 401   169

[[2]]
   drkreg
sex FALSE TRUE
  0 477   123
  1 400   170

[[3]]
   drkreg
sex FALSE TRUE
  0 480   120
  1 403   167

[[4]]
   drkreg
sex FALSE TRUE
  0 472   128
  1 409   161

[[5]]
   drkreg
sex FALSE TRUE
  0 479   121
  1 398   172

attr(,"call")
with.imputationList(all, table(sex, drkreg))
```

We can now fit a logistic regression model for trends over time in drinking:

```
> model1 <- with(all, glm(drkreg ~ wave * sex, family = binomial()))
> MIcombine(model1)

Multiple imputation results:
      with.imputationList(all, glm(drkreg ~ wave * sex, family = binomial()))
      MIcombine.default(model1)
                results         se
(Intercept) -2.25974358 0.26830731
wave         0.24055250 0.06587423
sex          0.64905222 0.34919264
wave:sex    -0.03725422 0.08609199

> summary(MIcombine(model1))

Multiple imputation results:
      with.imputationList(all, glm(drkreg ~ wave * sex, family = binomial()))
      MIcombine.default(model1)
                results         se      (lower     upper) missInfo
(Intercept) -2.25974358 0.26830731 -2.78584855 -1.7336386      4 %
wave         0.24055250 0.06587423  0.11092461  0.3701804     12 %
sex          0.64905222 0.34919264 -0.03537187  1.3334763      1 %
wave:sex    -0.03725422 0.08609199 -0.20623121  0.1317228      7 %
```

For model objects with `coef` and `vcov` methods the extraction of coefficients and variances is automatic, but `MIextract` can still be used:

```
> beta <- MIextract(model1, fun = coef)
> vars <- MIextract(model1, fun = vcov)
> summary(MIcombine(beta, vars))

Multiple imputation results:
      MIcombine.default(beta, vars)
                results         se      (lower     upper) missInfo
(Intercept) -2.25974358 0.26830731 -2.78584855 -1.7336386      4 %
wave         0.24055250 0.06587423  0.11092461  0.3701804     12 %
sex          0.64905222 0.34919264 -0.03537187  1.3334763      1 %
wave:sex    -0.03725422 0.08609199 -0.20623121  0.1317228      7 %
```

References

Carlin JB, Li N, Greenwood P, Coffey C. (2003) Tools for analyzing multiply imputed datasets. *Stata Journal* 3:1–20.