CHAPTER 11 ANALYSIS EXAMPLES REPLICATION - R SURVEY PACKAGE 3.22

GENERAL NOTES ABOUT ANALYSIS EXAMPLES REPLICATION

These examples are intended to provide guidance on how to use the commands/procedures for analysis of complex sample survey data and assume all data management and other preliminary work is done. The relevant syntax for the procedure of interest is shown first along with the associated output for that procedure(s). In some examples, there may be more than one block of syntax and in this case all syntax is first presented followed by the output produced.

In some software packages certain procedures or options are not available but we have made every attempt to demonstrate how to match the output produced by Stata 10+ in the textbook. Check the ASDA website for updates to the various software tools we cover.

NOTES ABOUT MULTIPLE IMPUTATION ANALYSES IN R SURVEY PACKAGE 3.22 (WITH R 2.7)

The R survey package used in these examples is 3.22 and was run under R 2.7 on a PC.

The R survey package offers a number of tools for analysis of multiply imputed data sets (imputed using SAS v9.2 in this case). These tools are easily combined with the various "svy" commands of R for correct analysis of survey data. The example provided uses the m=5 imputed NHANES data sets (previously imputed by SAS) stored as Stata type data sets in ".dta" format. These 5 data sets are first read into R and organized by imputation using the "imputationList" command. Once the data are read correctly, use of "svyglm" for survey linear regression is used to perform a separate linear regression for each of the imputed data sets. Next, use of the "MIcombine" command provides the analysis of the m=5 imputed data sets along with the survey corrections from "svyglm". Use of these commands with imputed data provides an easy way to correctly analyze imputed data based on a complex sample survey. See the R Survey documentation and general documentation for more details.
# read in data in .dta (stata) format stored in the mitools subdirectory under the library subdir

```r
> data.dir <- system.file("dta", package="mitools")
> library(foreign)
```

# read all files with c11_imp*.dta : there are 5 multiple imputed data sets stored

```r
> all <- imputationList(lapply(list.files(data.dir, pattern = "c11_imp.*.dta", full=TRUE),read.dta, warn.missing.labels=FALSE))
```

# examine the object "all"

```r
> all
```

MI data with 5 datasets

Call: imputationList(lapply(list.files(data.dir, pattern = "c11_imp.*.dta", full = TRUE), read.dta, warn.missing.labels = FALSE))

# check column names

```r
> colnames(all)
```

1. `_imputation_`
2. `riagendr`
3. `ridreth1`
4. `indfmpir`
5. `wtmec2yr`
6. `sdmvpus`
7. `sdmvstra`
8. `marcat`
9. `bpxdi1_1`
10. `age18p`
11. `agec`
12. `agecsq`
13. `bmxbmi`
14. `randval`
15. `married`
16. `swd`
17. `nevermarried`
18. `male`
19. `female`
20. `mex`
21. `othhis`
22. `white`
23. `black`
24. `other`

# declare survey variables and weight for svyglm analysis to come

```r
> des <- svydesign(id=~sdmvpus, strat=~sdmvstra, weight=~wtmec2yr, data=(all), nest=TRUE)

> des
```

Multiple (5) imputations: svydesign(id = ~sdmvpus, strat = ~sdmvstra, weight = ~wtmec2yr, data = (all), nest = TRUE)

# use svyglm for complex sample linear regression with 5 imputed data sets

# note that 5 sets of output are included here: [[1]] means first imputation output etc.

```r
> with (des, svyglm (bpxdi1_1 ~ othhis + white + black + other + swd + nevermarried + female + agec + agecsq + bmxbmi + indfmpir ))
```

[[1]]

Stratified 1 - level Cluster Sampling design (with replacement)

With (30) clusters.

Call: svyglm(bpxdi1_1 ~ othhis + white + black + other + swd + nevermarried + female + agec + agecsq + bmxbmi + indfmpir, design = .design)

Coefficients:

<table>
<thead>
<tr>
<th></th>
<th>othhis</th>
<th>white</th>
<th>black</th>
<th>other</th>
<th>swd</th>
<th>nevermarried</th>
<th>female</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intercept)</td>
<td>68.04061</td>
<td>1.08902</td>
<td>2.09394</td>
<td>2.44394</td>
<td>1.32406</td>
<td>0.74741</td>
<td>-0.34411</td>
</tr>
<tr>
<td>agec</td>
<td>0.10465</td>
<td>1.08902</td>
<td>2.09394</td>
<td>2.44394</td>
<td>1.32406</td>
<td>0.74741</td>
<td>-0.34411</td>
</tr>
<tr>
<td>agecsq</td>
<td>0.01028</td>
<td>0.18997</td>
<td>-0.02681</td>
<td>0.00000</td>
<td>0.00000</td>
<td>0.00000</td>
<td>0.00000</td>
</tr>
<tr>
<td>bmxbmi</td>
<td>-0.01028</td>
<td>0.18997</td>
<td>-0.02681</td>
<td>0.00000</td>
<td>0.00000</td>
<td>0.00000</td>
<td>0.00000</td>
</tr>
<tr>
<td>indfmpir</td>
<td>-0.00000</td>
<td>0.00000</td>
<td>0.00000</td>
<td>0.00000</td>
<td>0.00000</td>
<td>0.00000</td>
<td>0.00000</td>
</tr>
</tbody>
</table>

Degrees of Freedom: 5333 Total (i.e. Null); 4 Residual

Null Deviance: 820000
Residual Deviance: 720000 AIC: 43290
Stratified 1-level Cluster Sampling design (with replacement) With (30) clusters.

\[
\text{svydesign(ids = ids, probs = probs, strata = strata, variables = variables,}
\text{ fpc = fpc, nest = nest, check.strata = check.strata, weights = weights,}
\text{ data = d, pps = pps, ...})
\]

Call: \text{svyglm(bpxdi1_1 ~ othhis + white + black + other + swd + nevermarried +}
\text{ female + agec + agecsq + bmxbmi +}
\text{ indfmpir, design = .design)}

Coefficients:

\[
\begin{array}{cccccccc}
(\text{Intercept}) & \text{othhis} & \text{white} & \text{black} & \text{other} & \text{swd} & \text{nevermarried} & \text{female} \\
66.700445 & 1.631765 & 2.264346 & 3.721396 & 1.613467 & 0.519182 & -1.316616 & -2.847899 \\
0.103511 & 0.209438 & 0.050728 & \\
\end{array}
\]

Degrees of Freedom: 5333 Total (i.e. Null); 4 Residual
Null Deviance: 816300
Residual Deviance: 715100 AIC: 43260

Stratified 1-level Cluster Sampling design (with replacement) With (30) clusters.

\[
\text{svydesign(ids = ids, probs = probs, strata = strata, variables = variables,}
\text{ fpc = fpc, nest = nest, check.strata = check.strata, weights = weights,}
\text{ data = d, pps = pps, ...})
\]

Call: \text{svyglm(bpxdi1_1 ~ othhis + white + black + other + swd + nevermarried +}
\text{ female + agec + agecsq + bmxbmi +}
\text{ indfmpir, design = .design)}

Coefficients:

\[
\begin{array}{cccccccc}
(\text{Intercept}) & \text{othhis} & \text{white} & \text{black} & \text{other} & \text{swd} & \text{nevermarried} & \text{female} \\
67.79067 & 1.31006 & 2.03813 & 3.03001 & 3.09465 & 0.64236 & -0.19625 & -2.57246 \\
0.11474 & 0.17641 & 0.03998 & \\
\end{array}
\]

Degrees of Freedom: 5333 Total (i.e. Null); 4 Residual
Null Deviance: 813900
Residual Deviance: 715900 AIC: 43260

Stratified 1-level Cluster Sampling design (with replacement) With (30) clusters.

\[
\text{svydesign(ids = ids, probs = probs, strata = strata, variables = variables,}
\text{ fpc = fpc, nest = nest, check.strata = check.strata, weights = weights,}
\text{ data = d, pps = pps, ...})
\]

Call: \text{svyglm(bpxdi1_1 ~ othhis + white + black + other + swd + nevermarried +}
\text{ female + agec + agecsq + bmxbmi +}
\text{ indfmpir, design = .design)}

Coefficients:

\[
\begin{array}{cccccccc}
(\text{Intercept}) & \text{othhis} & \text{white} & \text{black} & \text{other} & \text{swd} & \text{nevermarried} & \text{female} \\
67.68423 & 0.80067 & 2.32955 & 3.58826 & 2.23942 & 0.49347 & -1.05222 & -3.01606 \\
0.09969 & 0.20085 & 0.10400 & \\
\end{array}
\]

Degrees of Freedom: 5333 Total (i.e. Null); 4 Residual
Null Deviance: 816500
Residual Deviance: 712300 AIC: 43240
Stratified 1-level Cluster Sampling design (with replacement)

With (30) clusters.

svydesign(ids = ids, probs = probs, strata = strata, variables = variables,
          fpc = fpc, nest = nest, check.strata = check.strata, weights = weights,
          data = d, pps = pps, ...)

Call: svyglm(bpxdi1_1 ~ othhis + white + black + other + swd + nevermarried +
           female + agec + agecsq + bmxbmi +
           indfmpir, design = .design)

Coefficients:
(Intercept)        othhis         white         black         other           swd  nevermarried        female
agec       66.95377       2.15371       2.29936       3.26213       2.44001       0.71647 -0.34554 -2.70136
0.11510
agecsq        0.01012       0.20505       0.06296

Degrees of Freedom: 5333 Total (i.e. Null); 4 Residual
Null Deviance: 812400
Residual Deviance: 713900  AIC: 43250

attr(,"call")
with(des, svyglm(bpxdi1_1 ~ othhis + white + black + other + swd + nevermarried +
           female + agec + agecsq + bmxbmi + indfmpir))

# Combine results for MI analysis of results (variance will be corrected due to imputation)
# save object called "c11imp" for use with MIcombine
> c11imp <- with (des, svyglm (bpxdi1_1 ~ othhis + white + black + other +
                           swd + nevermarried + female + agec + agecsq +
                           bmxbmi + indfmpir ))

> MIcombine(c11imp)
Multiple imputation results:
with(des, svyglm(bpxdi1_1 ~ othhis + white + black + other +
                 swd + nevermarried + female + agec + agecsq +
                 bmxbmi + indfmpir))
MIcombine.default(c11imp)

results  se
(Intercept)  67.43394438 1.349454913
othhis     1.39704366 1.143927741
white      2.0506439 0.584418299
black      3.20914736 0.886676870
other      0.14231942 1.083257937
swd         0.62377827 0.639336805
nevermarried -0.65094605 0.804572333
female     -2.86466982 0.432173702
agec       0.10753883 0.015780413
agecsq     -0.01015275 0.000898504
bmxbmi     0.19834482 0.036645441
indfmpir    0.02061374 0.147957736
# use of summary option for more information from MIcombine
> summary(MIcombine(c11imp))

Multiple imputation results:

```r
with(des, svyglm(bpxdi1_1 ~ othhis + white + black + other +
    swd + nevermarried + female + agec + agecsq + bmxbmi + indfmpir))
MIcombine.default(c11imp)
```

<table>
<thead>
<tr>
<th></th>
<th>results</th>
<th>se</th>
<th>(lower</th>
<th>upper</th>
<th>missInfo</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intercept)</td>
<td>67.43394438</td>
<td>1.349454913</td>
<td>64.75029115</td>
<td>70.117597613</td>
<td>24 %</td>
</tr>
<tr>
<td>othhis</td>
<td>1.39704366</td>
<td>1.143927741</td>
<td>-0.88783344</td>
<td>3.681920764</td>
<td>27 %</td>
</tr>
<tr>
<td>white</td>
<td>2.20506439</td>
<td>0.584418299</td>
<td>1.05838340</td>
<td>3.351745369</td>
<td>6 %</td>
</tr>
<tr>
<td>black</td>
<td>3.20914736</td>
<td>0.886676870</td>
<td>1.38678653</td>
<td>5.031508193</td>
<td>43 %</td>
</tr>
<tr>
<td>other</td>
<td>2.14231942</td>
<td>1.130257937</td>
<td>-0.22339884</td>
<td>4.508037679</td>
<td>51 %</td>
</tr>
<tr>
<td>swd</td>
<td>0.62377827</td>
<td>0.639336805</td>
<td>-0.62985438</td>
<td>1.877410925</td>
<td>4 %</td>
</tr>
<tr>
<td>nevermarried</td>
<td>-0.65094605</td>
<td>0.804572333</td>
<td>-2.33684465</td>
<td>1.034952556</td>
<td>51 %</td>
</tr>
<tr>
<td>female</td>
<td>-2.86466982</td>
<td>0.432173702</td>
<td>-3.75093409</td>
<td>-1.978405550</td>
<td>42 %</td>
</tr>
<tr>
<td>agec</td>
<td>0.10753883</td>
<td>0.015780413</td>
<td>0.07608372</td>
<td>0.138993946</td>
<td>26 %</td>
</tr>
<tr>
<td>agecsq</td>
<td>-0.01015275</td>
<td>0.000898504</td>
<td>-0.01192643</td>
<td>-0.008379067</td>
<td>16 %</td>
</tr>
<tr>
<td>bmxbmi</td>
<td>0.19634482</td>
<td>0.036645441</td>
<td>0.12397704</td>
<td>0.268712610</td>
<td>17 %</td>
</tr>
<tr>
<td>indfmpir</td>
<td>-0.02061374</td>
<td>0.147957736</td>
<td>-0.31576716</td>
<td>0.274539685</td>
<td>26 %</td>
</tr>
</tbody>
</table>