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

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

# read all files with c11_imp*.dta : there are 5 multiple imputed data sets stored
> all <- imputationList(lapply(list.files(data.dir, pattern = "c11_imp.\\.dta", full=TRUE),read.dta,
warn.missing.labels=FALSE))

# examine the object "all"
> 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
> colnames(all)
 [1] "_imputation_" "riagendr"    "ridreth1"    "indfmpir"    "wtmec2yr"    "sdmvpsu"     "sdmvstra"    "marcat"
 [9] "bpxdi1_1"     "age18p"      "agec"        "agecsq"      "bmxbmi"      "randval"     "married"     "swd"
[17] "nevermarried" "male"        "female"      "mex"         "othhis"      "white"       "black"       "other"

# declare survey variables and weight for svyglm analysis to come
> des <- svydesign(id=~sdmvpsu, strat=~sdmvstra, weight=~wtmec2yr, data=(all), nest=TRUE)

> des
Multiple (5) imputations: svydesign(id = ~sdmvpsu, 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.

> 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.
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
    68.04061       1.08902       2.09394       2.44394       1.32406       0.74741       -0.34411       -3.18557
0.10465
      agecsq        bmxbmi        indfmpir
    -0.01028       0.18997       -0.02681

Degrees of Freedom: 5333 Total (i.e. Null);  4 Residual
Null Deviance:      820000
Residual Deviance: 720000       AIC: 43290
```

```
[[2]]
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
   66.700445      1.631765      2.264346      3.721396      1.613467      0.519182     -1.316616     -2.847899
0.103511
       agecsq        bmxbmi       indfmpir
    -0.009621      0.209438      0.050728

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

[[3]]
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
    67.79067       1.31006       2.03813       3.03001       3.09465       0.64236      -0.19625      -2.57246
0.11474
       agecsq        bmxbmi       indfmpir
     -0.01046       0.17641       0.03998

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

[[4]]
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
    67.68423       0.80067       2.32955       3.58826       2.23942       0.49347      -1.05222      -3.01606
0.09969
       agecsq        bmxbmi       indfmpir
     -0.01028       0.20085      -0.10400

Degrees of Freedom: 5333 Total (i.e. Null);  4 Residual
Null Deviance:       816500
Residual Deviance: 712300       AIC: 43240
```

```
[[5]]
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          bmxbmi         indfmpir
     -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         2.20506439 0.584418299
black         3.20914736 0.886676870
other         2.14231942 1.130257937
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        0.19634482 0.036645441
indfmpir     -0.02061374 0.147957736
```

```
# use of summary option for more information from MIcombine
> summary(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      (lower        upper) missInfo
(Intercept)  67.43394438 1.349454913 64.75029115 70.117597613    24 %
othhis        1.39704366 1.143927741 -0.88783344  3.681920764    27 %
white         2.20506439 0.584418299  1.05838340  3.351745369     6 %
black         3.20914736 0.886676870  1.38678653  5.031508193    43 %
other         2.14231942 1.130257937 -0.22339884  4.508037679    51 %
swd           0.62377827 0.639336805 -0.62985438  1.877410925     4 %
nevermarried -0.65094605 0.804572333 -2.33684465  1.034952556    51 %
female       -2.86466982 0.432173702 -3.75093409 -1.978405550    42 %
agec          0.10753883 0.015780413  0.07608372  0.138993946    26 %
agecsq       -0.01015275 0.000898504 -0.01192643 -0.008379067    16 %
bmxbmi        0.19634482 0.036645441  0.12397704  0.268712610    17 %
indfmpir     -0.02061374 0.147957736 -0.31576716  0.274539685    26 %
```