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.

GENERAL NOTES ABOUT CHAPTER 8 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 very good range of svy commands for the analyses of this chapter: logistic, probit, and Clog-log models are all available via the svyglm command with the correct family/link options. Other features included in this chapter are use of a factor statement for categorical variables, examples of how to obtain odds ratios from an object created from a model command, use of the regTermTest command for testing of groups of parameters including interactions in models, and bivariate testing of predictors for models using the svychisq command.
# Data production and set up of design objects
# remember to load package first survey package

### NHANES

```r
nhanesdata <- read.table(file = "f:/applied_analysis_book/r/nhanes_final.txt", sep = "\t", header = T, as.is=T)

#create factor variables
nhanesdata$racec <- factor(nhanesdata$RIDRETH1, levels = 1: 5 , labels =c("Mexican", "Other Hispanic", "White", "Black", "Other"))
nhanesdata$marcatc <- factor(nhanesdata$marcat, levels = 1: 3, labels =c("Married", "Previously Married", "Never Married"))
nhanesdata$edcatc <- factor(nhanesdata$edcat, levels = 1: 4, labels =c("0-11", "12", "13-15","16+"))
nhanesdata$bp_catc <- factor(nhanesdata$bp_cat, levels = 1: 4, labels =c("Normal", "Pre-HBP", "Stage 1 HBP","Stage 2 HBP"))
nhanesdata$agesq <- (nhanesdata$agecent * nhanesdata$agecent )
```

```r
names(nhanesdata)
```

```r
nhanessvy2 <- svydesign(strata=~SDMVSTRA, id=~SDMVPSU, weights=~WTMEC2YR, data=nhanesdata, nest=T)
subnhanes <- subset(nhanessvy2 , RIDAGEYR >= 18)
```

### NCS-R

```r
ncsr <- read.table(file = "f:/applied_analysis_book/r/ncsr2010.txt", sep = "\t", header = T, as.is=T)
```

```r
# create factor versions with labels
ncsr$racec <- factor(ncsr$racecat, levels = 1: 4, labels =c("Other", "Hispanic", "Black", "White"))
n CSR$marcatc <- factor(ncsr$MAR3CAT, levels = 1: 3, labels =c("Married", "Previously Married", "Never Married"))
n CSR$edcatc <- factor(ncsr$ED4CAT, levels = 1: 4, labels =c("0-11", "12", "13-15","16+"))
n CSR$sexc <- factor(ncsr$SEX, levels = 1:2, labels=c("Male","Female"))
n CSR$agcatc <- factor(ncsr$ag4cat, levels = 1:4, labels=c("18-29", "30-44", "45-59", "60+"))
```

```r
ncsrsvyp1 <- svydesign(strata=~SESTRAT, id=~SECLUSTR, weights=~NCSRWTSH, data=ncsr, nest=T)
ncsrsvyp2 <- svydesign(strata=~SESTRAT, id=~SECLUSTR, weights=~NCSRWTLG, data=ncsr, nest=T)
ncsrsvypop <- svydesign(strata=~SESTRAT, id=~SECLUSTR, weights=~popweight, data=ncsr, nest=T)
```

### HRS

# both hh and r weights are needed plus financial respondent for hh level analysis
```r
hrs <- read.table(file = "f:/applied_analysis_book/r/hrs2010.txt", sep = "\t", header = T, as.is=T)
hrssvyhh <- svydesign(strata=~STRATUM, id=~SECU, weights=~KWGTHH , data=hrs, nest=T)
summary(hrssvyhh)
```

```r
hrssvyhh <- subset(hrssvyhh, KFINR==1)
```

```r
hrssvyr <- svydesign(strata=~STRATUM, id=~SECU, weights=~KWGTR , data=hrs, nest=T)
summary(hrssvyr)
```
#EXAMPLE 8.1 BIVARIATE TESTING OF PREDICTORS OF MDE NCS-R DATA

#USE OF SVYCHISQ TO OBTAIN 2 BY 2 CHISQ TESTS OF ASSOCIATION

```r
> svychisq(~mde + ag4cat, ncsrsvyp2)

Pearson's X^2: Rao & Scott adjustment

data:  svychisq(~mde + ag4cat, ncsrsvyp2)
F = 26.3902, ndf = 2.761, ddf = 115.970, p-value = 1.965e-12

> svychisq(~mde + SEX, ncsrsvyp2)

Pearson's X^2: Rao & Scott adjustment

data:  svychisq(~mde + SEX, ncsrsvyp2)
F = 44.8339, ndf = 1, ddf = 42, p-value = 3.965e-08

> svychisq(~mde + ald, ncsrsvyp2)

Pearson's X^2: Rao & Scott adjustment

data:  svychisq(~mde + ald, ncsrsvyp2)
F = 120.0282, ndf = 1, ddf = 42, p-value = 6.86e-13

> svychisq(~mde + ED4CAT, ncsrsvyp2)

Pearson's X^2: Rao & Scott adjustment

data:  svychisq(~mde + ED4CAT, ncsrsvyp2)
F = 4.3043, ndf = 2.903, ddf = 121.925, p-value = 0.006916

> svychisq(~mde + MAR3CAT, ncsrsvyp2)

Pearson's X^2: Rao & Scott adjustment

data:  svychisq(~mde + MAR3CAT, ncsrsvyp2)
F = 11.0849, ndf = 1.899, ddf = 79.745, p-value = 7.616e-05
```
# CHAPTER 8 ANALYSIS EXAMPLE 8.1: NCSR DATA

```r
> #model 8.1
> ex81 <- svyglm(mde ~ factor(agcatc) + sexm + ald + factor(edcatc) + factor(marcatc), family=quasibinomial, design=ncsrsvyp2)

Warning messages:
1: In summary.glm(glm) :
  observations with zero weight not used for calculating dispersion
2: In summary.glm(glm.object) :
  observations with zero weight not used for calculating dispersion
> summary(ex81)

Call:
svyglm(mde ~ factor(agcatc) + sexm + ald + factor(edcatc) + factor(marcatc),
      family = quasibinomial, design = ncsrsvyp2)

Survey design:
svydesign(strata = ~SESTRAT, id = ~SECLUSTR, weights = ~NCSRWTLG,
           data = ncsr, nest = T)

Coefficients:                Estimate Std. Error t value Pr(>|t|)
(Intercept)                   -1.58308    0.12066  -13.120  2.01e-14 ***
factor(agcatc)30-44          0.25562    0.09438   2.708   0.0108 *
factor(agcatc)45-59          0.20645    0.09153   2.256   0.0311 *
factor(agcatc)60+            -0.67579    0.14130  -4.783  3.74e-05 ***
sexm                          0.57735    0.07722   7.477  1.64e-10 ***
ald                           0.42368    0.07602   5.588  3.69e-08 ***
factor(edcatc)12              0.07925    0.09690   0.818   0.4194
factor(edcatc)13-15           0.23051    0.09307   2.477   0.0187 *
factor(edcatc)16+             0.16293    0.11061   1.473   0.1505
factor(marcatc)Previously Married 0.48642    0.10542   4.605  3.60e-09 ***
factor(marcatc)Never Married  0.11558    0.10787   1.071   0.2920

---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for quasibinomial family taken to be 1.001496)

Number of Fisher Scoring iterations: 5
```

#exp of coefficients obtained by use of the exp command with the ex81 coefficients

```r
> #exp of coefficients
> exp(ex81$coef)

(Intercept)                factor(agcatc)30-44
  0.2053424                   1.2912600
factor(agcatc)45-59 factor(agcatc)60+
  1.2293019                   0.5087563
sexm                        ald
  0.5613867                   4.1523575
factor(edcatc)12 factor(edcatc)13-15
  1.082403                    1.2592434
factor(edcatc)16+ factor(marcatc)Previously Married
  5.695 2.63e-06
factor(marcatc)Never Married
  1.1769489                   1.6264870
factor(marcatc)Previously Married 0.48642    0.10542   4.605  3.60e-09 ***
factor(marcatc)Never Married  0.11558    0.10787   1.071   0.2920
```
> # tests of parameters
> regTermTest(ex81, ~factor(agcatc))
> Wald test for factor(agcatc)
> in svyglm(mde ~ factor(agcatc) + sexm + ald + factor(edcatc) + factor(marcatc),
> family = quasibinomial, design = ncsrsvyp2)
> Chisq = 59.94856 on 3 df: p = 6.0289e-13

> regTermTest(ex81, ~factor(edcatc))
> Wald test for factor(edcatc)
> in svyglm(mde ~ factor(agcatc) + sexm + ald + factor(edcatc) + factor(marcatc),
> family = quasibinomial, design = ncsrsvyp2)
> Chisq = 6.708987 on 3 df: p = 0.081775

> regTermTest(ex81, ~factor(marcatc))
> Wald test for factor(marcatc)
> in svyglm(mde ~ factor(agcatc) + sexm + ald + factor(edcatc) + factor(marcatc),
> family = quasibinomial, design = ncsrsvyp2)
> Chisq = 34.0155 on 2 df: p = 4.108e-08

# NO SVY GOF TEST AVAILABLE IN R SURVEY PACKAGE
# EXAMPLE 8.1 NCS-R DATA INTERACTION TESTS OF SEX * EACH OTHER PREDICTOR IN MODEL

> # add interactions to model

> ex81_int <- svyglm(mde ~ factor(agcatc) + sexm + ald + factor(edcatc) + factor(marcatc) + sexm*factor(agcatc) + sexm*ald + sexm*factor(edcatc) + sexm*factor(marcatc), family=quasibinomial, design=ncsrsvyp2)

Warning messages:
1: In summary.glm(g) :
   observations with zero weight not used for calculating dispersion
2: In summary.glm(glm.object) :
   observations with zero weight not used for calculating dispersion

> summary(ex81_int)

Call:
svyglm(mde ~ factor(agcatc) + sexm + ald + factor(edcatc) + factor(marcatc) +
       sexm * factor(agcatc) + sexm * ald + sexm * factor(edcatc) +
       sexm * factor(marcatc), family = quasibinomial, design = ncsrsvyp2)

Survey design:
svydesign(strata = ~SESTRAT, id = ~SECLUSTR, weights = ~NCSRWTLG,
data = ncsr, nest = T)

Coefficients:

                      Estimate Std. Error t value  Pr(>|t|)
(Intercept)          -1.599889   0.134010 -11.939   9.25e-11 ***
factor(agcatc)30-44    0.220404   0.113761   1.937    0.06507 .
factor(agcatc)45-59    0.214641   0.102494   2.094    0.04746 *
factor(agcatc)60+      -0.645556   0.175192  -3.685    0.00123 **
sexm                   -0.546442   0.357165  -1.530    0.13967
ald                     1.553140   0.211023   7.360   1.74e-07 ***
factor(edcatc)12       0.130518   0.083710   1.559    0.13261
factor(edcatc)13-15    0.297324   0.117042   2.540    0.01287 *
factor(edcatc)16+      0.242218   0.151870   1.595    0.12438
factor(marcatc)Previously Married 0.417786   0.110525   3.780    0.00097 ***
factor(marcatc)Never Married      0.017337   0.129782   0.134    0.89489
factor(agcatc)30-44:sexm  0.096743   0.200841   0.482    0.63458
factor(agcatc)45-59:sexm 0.002637   0.212819   0.012    0.99022
factor(agcatc)60+:sexm  -0.037810   0.302039  -0.125    0.90147
sexm:ald                -0.200417   0.242244  -0.827    0.41665
sexm:factor(edcatc)12   -0.137780   0.271016  -0.508    0.61603
sexm:factor(edcatc)13-15 -0.166790   0.269345  -0.627    0.53705
sexm:factor(edcatc)16+   -0.194018   0.344168  -0.564    0.57839
sexm:factor(marcatc)Previously Married 0.182504   0.207876   0.878    0.38905
sexm:factor(marcatc)Never Married      0.231898   0.211962   1.094    0.28526

---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for quasibinomial family taken to be 1.001791)

Number of Fisher Scoring iterations: 5
> regTermTest(ex81_int, ~sexm:factor(agcatc))
Wald test for sexm:factor(agcatc)
  in svyglm(mde ~ factor(agcatc) + sexm + ald + factor(edcatc) + factor(marcatc) +
  sexm * factor(agcatc) + sexm * ald + sexm * factor(edcatc) +
  sexm * factor(marcatc), family = quasibinomial, design = ncsrsvyp2)
  Chisq = 0.7799 on 3 df: p = 0.85427

> regTermTest(ex81_int, ~sexm:ald)
Wald test for sexm:ald
  in svyglm(mde ~ factor(agcatc) + sexm + ald + factor(edcatc) + factor(marcatc) +
  sexm * factor(agcatc) + sexm * ald + sexm * factor(edcatc) +
  sexm * factor(marcatc), family = quasibinomial, design = ncsrsvyp2)
  Chisq = 0.6844827 on 1 df: p = 0.40805

> regTermTest(ex81_int, ~sexm:factor(edcatc))
Wald test for sexm:factor(edcatc)
  in svyglm(mde ~ factor(agcatc) + sexm + ald + factor(edcatc) + factor(marcatc) +
  sexm * factor(agcatc) + sexm * ald + sexm * factor(edcatc) +
  sexm * factor(marcatc), family = quasibinomial, design = ncsrsvyp2)
  Chisq = 0.3954802 on 3 df: p = 0.94117

> regTermTest(ex81_int, ~sexm:factor(marcatc))
Wald test for sexm:factor(marcatc)
  in svyglm(mde ~ factor(agcatc) + sexm + ald + factor(edcatc) + factor(marcatc) +
  sexm * factor(agcatc) + sexm * ald + sexm * factor(edcatc) +
  sexm * factor(marcatc), family = quasibinomial, design = ncsrsvyp2)
  Chisq = 1.567496 on 2 df: p = 0.45669
# EXAMPLE 8.2 NCS-R DATA: COMPARISON OF LOGIT, PROBIT AND CLOGLOG

# EXAMPLE 8.2: LOGIT

> # model 8.2 with logit probit and cloglog comparison

> summary(ex82_logit <- svyglm(ald ~ factor(agcatc) + sexm + factor(edcatc) + factor(marcatc),
          family=quasibinomial, design=ncsrsvyp2))

Call:
svyglm(ald ~ factor(agcatc) + sexm + factor(edcatc) + factor(marcatc),
      family = quasibinomial, design = ncsrvsvp2)

Survey design:
svydesign(strata = ~SESTRAT, id = ~SECLUSTR, weights = ~NCSRWTLG,
          data = ncsr, nest = T)

Coefficients:                        Estimate Std. Error t value Pr(>|t|)
(Intercept)                       -3.12432    0.22527 -13.869 2.54e-15 ***
factor(agcatc)30-44                 0.14628    0.17813   0.821 0.417447
factor(agcatc)45-59                -0.05071    0.14392  -0.352 0.726830
factor(agcatc)60+                  -1.12033    0.21247  -5.273 8.25e-06 ***
sexm                               0.99799    0.11912   8.378 1.12e-09 ***
factor(edcatc)12                   -0.26844    0.19373  -1.386 0.175157
factor(edcatc)13-15               -0.26448    0.17614  -1.502 0.142722
factor(edcatc)16+                  -0.73623    0.19719  -3.734 0.000712 ***
factor(marcatc)Previously Married  0.51783    0.14207   3.645 0.000910 ***
factor(marcatc)Never Married       0.06532    0.16870   0.387 0.701121
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for quasibinomial family taken to be 0.9750455)

Number of Fisher Scoring iterations: 6

Warning messages:
1: In summary.glm(g) :
   observations with zero weight not used for calculating dispersion
2: In summary.glm(glm.object) :
   observations with zero weight not used for calculating dispersion

> regTermTest(ex82_logit, ~factor(agcatc))
Wald test for factor(agcatc)
  in svyglm(ald ~ factor(agcatc) + sexm + factor(edcatc) + factor(marcatc),
    family = quasibinomial, design = ncsrvsvp2)
  Chisq =  37.98918  on  3  df: p= 2.8414e-08

> regTermTest(ex82_logit, ~factor(edcatc))
Wald test for factor(edcatc)
  in svyglm(ald ~ factor(agcatc) + sexm + factor(edcatc) + factor(marcatc),
    family = quasibinomial, design = ncsrvsvp2)
  Chisq =  15.11116  on  3  df: p= 0.0017241

> regTermTest(ex82_logit, ~factor(marcatc))
Wald test for factor(marcatc)
  in svyglm(ald ~ factor(agcatc) + sexm + factor(edcatc) + factor(marcatc),
    family = quasibinomial, design = ncsrvsvp2)
  Chisq =  13.39217  on  2  df: p= 0.0012357
#EXAMPLE 8.2: PROBIT

> summary(ex82_probit <- svyglm(ald ~ factor(agcatc) + sexm + factor(edcatc) + factor(marcatc),
family=quasibinomial(link=probit), design=ncrsrvyp2))

Call:
svyglm(ald ~ factor(agcatc) + sexm + factor(edcatc) + factor(marcatc),
family = quasibinomial(link = probit), design = ncrsrvyp2)

Survey design:
svydesign(strata = ~SESTRAT, id = ~SECLUSTR, weights = ~NCSRWTLG,
data = ncsr, nest = T)

Coefficients:

|              | Estimate | Std. Error | t value | Pr(>|t|) |
|--------------|----------|------------|---------|----------|
| (Intercept)  | -1.7194  | 0.1064     | -16.149 | < 2e-16  *** |
| factor(agcatc)30-44 | 0.0653 | 0.0854 | 0.765   | 0.4498        |
| factor(agcatc)45-59 | -0.0344 | 0.0673 | -0.512  | 0.6121        |
| factor(agcatc)60+  | -0.5312 | 0.0937 | -5.669  | 2.56e-06     *** |
| sexm          | 0.4708  | 0.0569     | 8.267   | 1.51e-09     *** |
| factor(edcatc)12 | -0.1237 | 0.0949 | -1.303  | 0.2014        |
| factor(edcatc)13-15 | -0.1243 | 0.0850 | -1.463  | 0.1530        |
| factor(edcatc)16+ | -0.3396 | 0.0924 | -3.676  | 0.000835      *** |
| factor(marcatc)Previously Married | 0.2547 | 0.0701 | 3.633   | 0.000942      *** |
| factor(marcatc)Never Married | 0.0389 | 0.0779 | 0.500   | 0.6206        |

---

Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for quasibinomial family taken to be 0.9824716)

Number of Fisher Scoring iterations: 6

Warning messages:
1: In summary.glm(g) :
  observations with zero weight not used for calculating dispersion
2: In summary.glm(glm.object) :
  observations with zero weight not used for calculating dispersion

> regTermTest(ex82_probit, ~factor(agcatc))

Wald test for factor(agcatc)
in svyglm(ald ~ factor(agcatc) + sexm + factor(edcatc) + factor(marcatc),
family = quasibinomial(link = probit), design = ncrsrvyp2)
Chisq = 48.0261 on 3 df: p= 2.1024e-10

> regTermTest(ex82_probit, ~factor(edcatc))

Wald test for factor(edcatc)
in svyglm(ald ~ factor(agcatc) + sexm + factor(edcatc) + factor(marcatc),
family = quasibinomial(link = probit), design = ncrsrvyp2)
Chisq = 15.01579 on 3 df: p= 0.0018032

> regTermTest(ex82_probit, ~factor(marcatc))

Wald test for factor(marcatc)
in svyglm(ald ~ factor(agcatc) + sexm + factor(edcatc) + factor(marcatc),
family = quasibinomial(link = probit), design = ncrsrvyp2)
Chisq = 13.40266 on 2 df: p= 0.0012293
# EXAMPLE 8.2 CLOGLOG

```r
> summary(ex82_cloglog <- svyglm(ald ~ factor(agcatc) + sexm + factor(edcatc) + factor(marcatc), 
family=quasibinomial(link=cloglog), design=ncsrsvyp2))

Call:
svyglm(ald ~ factor(agcatc) + sexm + factor(edcatc) + factor(marcatc), 
family = quasibinomial(link = cloglog), design = ncsrsvyp2)

Survey design:
svydesign(strata = ~SESTRAT, id = ~SECLUSTR, weights = ~NCSRWTLG, 
data = ncsr, nest = T)

Coefficients:                        Estimate  Std. Error t value Pr(>|t|)
(Intercept)         -3.1484 0.21716   -14.498  7.18e-16 ***
factor(agcatc)30-44  0.14304 0.17125    0.835   0.409554
factor(agcatc)45-59 -0.04524 0.13967  -0.324    0.748043
factor(agcatc)60+    1.08287 0.20817    5.202   1.02e-05 ***
sexm                  0.96522 0.11494     8.398   1.06e-09 ***
factor(edcatc)12     -0.26010 0.18467   -1.408   0.168356
factor(edcatc)13-15  0.25556 0.16843     1.517   0.138698
factor(edcatc)16+    -0.71265 0.19042    -3.743  0.000694 ***
factor(marcatc)Previously Married 0.49354 0.13522     3.650  0.000898 ***
factor(marcatc)Never Married          0.06049 0.16369  0.370   0.714075
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 . ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for quasibinomial family taken to be 0.9745215)

Number of Fisher Scoring iterations: 7

Warning messages:
1: In summary.glm(g) :
  observations with zero weight not used for calculating dispersion
2: In summary.glm(glm.object) :
  observations with zero weight not used for calculating dispersion
```

Warning messages:
1: In summary.glm(g) :
  observations with zero weight not used for calculating dispersion
2: In summary.glm(glm.object) :
  observations with zero weight not used for calculating dispersion

> regTermTest(ex82_cloglog, ~factor(agcatc))
Wald test for factor(agcatc)
in svyglm(ald ~ factor(agcatc) + sexm + factor(edcatc) + factor(marcatc), 
  family = quasibinomial(link = cloglog), design = ncsrsvyp2)
Chisq = 36.56382 on 3 df: p = 5.6906e-08

> regTermTest(ex82_cloglog, ~factor(edcatc))
Wald test for factor(edcatc)
in svyglm(ald ~ factor(agcatc) + sexm + factor(edcatc) + factor(marcatc), 
  family = quasibinomial(link = cloglog), design = ncsrsvyp2)
Chisq = 15.05531 on 3 df: p = 0.00177

> regTermTest(ex82_cloglog, ~factor(marcatc))
Wald test for factor(marcatc)
in svyglm(ald ~ factor(agcatc) + sexm + factor(edcatc) + factor(marcatc), 
  family = quasibinomial(link = cloglog), design = ncsrsvyp2)
Chisq = 13.42122 on 2 df: p = 0.0012179