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 6 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: svyby, svymean,
svychisq, svycontrast, and svyglm (for simple logistic regression) are all used to analyze categorical data. In
addition, use of the barplot and pie commands for graphics are demonstrated with output objects from the svy
commands. Other features included are use of a factor statement for categorical variables, the confint command
for confidence intervals, and the svychisq command for tests of association for two way tables.
# Data production and set up of design objects
# remember to load package first survey package

# NHANES
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)

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

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

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

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
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)
hrssvysub <- subset(hrssvyhh, KFINR==1)

hrssvyr <- svydesign(strata=~STRATUM, id=~SECU, weights=~KWGTR , data=hrs, nest=T)
summary(hrssvyr)
# CHAPTER 6 ANALYSIS EXAMPLES

# EXAMPLE OF CREATING FACTOR VARIABLES PRIOR TO USE WITH NHANES DATA
# FACTOR VARIABLES ARE USED AS LABELLED VARIABLES FOR EASY READING OF OUTPUT

```r
code
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$edca <- factor(nhanesdata$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"))
```

# ANALYSIS EXAMPLE 6.1 NHANES ADULT DATA

```r
> (ex61 <- svymean(~factor(irregular), subnhanes, se=T, na.rm=T, deff=T, ci=T, keep.vars=T))
  mean        SE    DEff
factor(irregular)0  0.9704690 0.0066546 7.9116
factor(irregular)1  0.0295310 0.0066546 7.9116
> confint(ex61)
2.5 %     97.5 %
factor(irregular)0  0.95742616 0.98351177
factor(irregular)1  0.01648823 0.04257384
```

# ANALYSIS EXAMPLE 6.2 NHANES ADULT DATA

```r
> (ex62 <- svymean(~factor(racec), subnhanes, se=T, na.rm=T, deff=T, ci=T, keep.vars=T))
  mean        SE    DEff
factor(racec)Mexican  0.0807834 0.0100534  7.5706
factor(racec)Other Hispanic  0.0337852 0.0074214  9.3845
factor(racec)White    0.7141428 0.0276981 20.9030
factor(racec)Black    0.1172628 0.0198491 21.1705
factor(racec)Other    0.0540257 0.0058250  3.6928
> confint(ex62)
2.5 %     97.5 %
factor(racec)Mexican  0.06107905 0.10048772
factor(racec)Other Hispanic  0.01923958 0.04833087
factor(racec)White    0.65985549 0.76843019
factor(racec)Black    0.07835937 0.15616630
factor(racec)Other    0.04260888 0.06544254
```

# ANALYSIS EXAMPLE 6.3 NHANES ADULT DATA

```r
> (ex63 <- svymean(~factor(bp_catc), subnhanes, se=T, na.rm=T, deff=T, ci=T, keep.vars=T))
  mean        SE    DEff
factor(bp_catc)Normal  0.4710800 0.0110853 2.4936
factor(bp_catc)Pre-HBP  0.4185420 0.0117868 2.8864
factor(bp_catc)Stage 1 HBP  0.0864090 0.0062077 2.4681
factor(bp_catc)Stage 2 HBP  0.0239689 0.0024045 1.2495
> confint(ex63)
2.5 %     97.5 %
factor(bp_catc)Normal  0.44935325 0.49280685
factor(bp_catc)Pre-HBP  0.39544028 0.44164378
factor(bp_catc)Stage 1 HBP  0.07424219 0.09857590
factor(bp_catc)Stage 2 HBP  0.01925624 0.02868152
```

# ANALYSIS EXAMPLE 6.4 USES THE GOF TOOL WITH PRE-SET PROPORTIONS NOT INCLUDED HERE
# ANALYSIS EXAMPLE 6.5 PIE AND BAR CHARTS

# PIE CHART OF BLOOD PRESSURE STATUS NHANES DATA

```r
(ex63 <- svymean(~factor(bp_cat), subnhanes, se=T, na.rm=T, deff=T, ci=T, keep.vars=T))

> pie(ex63, col=c("black", "grey60", "blue", "red"), c("Normal", "Pre-HBP", "Stage 1 HBP", "Stage 2 HBP"))
```
# BAR CHART OF BLOOD PRESSURE STATUS

```r
barplot(ex63, legend=c("Normal", "Pre-HBP", "Stage 1 HBP", "Stage 2 HBP"), col=c("black", "grey60", "blue", "red"))
```
# ANALYSIS EXAMPLE 6.6 NCS-R DATA

```r
> (ex66 <- svymean(~interaction(SEX, mde), ncsrsvyp1, se=T, na.rm=T, ci=T, keep.vars=T))

  mean SE
interaction(SEX, mde)1.0 0.406644 0.0070
interaction(SEX, mde)2.0 0.401644 0.0054
interaction(SEX, mde)1.1 0.072208 0.0034
interaction(SEX, mde)2.1 0.119504 0.0030

# obtain confidence intervals
> confint(ex66)

   2.5 %        97.5 %
interaction(SEX, mde)1.0 0.39296383 0.42032513
interaction(SEX, mde)2.0 0.39113771 0.41215085
interaction(SEX, mde)1.1 0.06546993 0.07894551
interaction(SEX, mde)2.1 0.11356911 0.12543793
```

# svyby analysis gives mean of mde by sex

**CODES FOR SEX 1=MALE 2=FEMALE**

```r
> (ex66 <- svyby(~mde, ~SEX, ncsrsvyp1, svymean, se=T, na.rm=T, ci=T, keep.vars=T))

      SEX       mde      se.mde
1   Male 0.1507933 0.007747811
2 Female 0.2293083 0.005647255
```

#svychisq provides a 2 by 2 chisq test (F)

```r
> svychisq(~SEX + mde, design = ncsrsvyp1, statistic = "F")

Pearson's X^2: Rao & Scott adjustment
data:  svychisq(~SEX + mde, design = ncsrsvyp1, statistic = "F")
F = 57.9784, ndf = 1, ddf = 42, p-value = 1.947e-09
```

# ANALYSIS EXAMPLE 6.7 MEAN OF MDE OVER SEX AND LINEAR COMPARISON TEST

**USE OF FACTOR VARIABLE WITH LABELS FOR SEX 1=MALE 2=FEMALE**

```r
> (ex66 <- svyby(~mde, ~sexc, ncsrsvyp1, svymean, se=T, na.rm=T, ci=T, keep.vars=T))

   sexc     mde      se.mde
Male     Male 0.1507933 0.007747811
Female Female 0.2293083 0.005647255
```

```r
> svycontrast(ex66,list(avg=c(.5,.5), diff=c(1,-1)))

  contrast     SE
avg 0.190051 0.0048
diff -0.078515 0.0096
Warning message:
In vcov.svyby(stat) : Only diagonal elements of vcov() available
```
# ANALYSIS EXAMPLE 6.8 NCS-R DATA WITH PART 2 WEIGHT

```r
ex68 <- svyby (~ald, ~edcatc, subset(ncsrsvyp2, AGE < 29 & !is.na(ED4CAT) & !is.na(ald)), svymean, na.rm=T, ci=T)
There were 16 warnings (use warnings() to see them)
```

#NOTE: WARNINGS PERTAIN THE LONELY PSU'S, SEE DOCUMENTATION FOR DETAILS ON HOW R HANDLES THIS SITUATION

```r
> print(ex68)
edcatc        ald     se.ald
0-11    0-11 0.09128575 0.02937999
12     12 0.04855850 0.01345971
13-15  13-15 0.04895775 0.01004206
16+    16+ 0.06903765 0.01364029
```

```r
> svychisq(~ald + edcatc, subset(ncsrsvyp2, AGE < 29 & !is.na(ED4CAT) & !is.na(ald)), statistic = "F")
Pearson's X^2: Rao & Scott adjustment
```

```r
data:  svychisq(~ald + edcatc, subset(ncsrsvyp2, AGE < 29 & !is.na(ED4CAT) & !is.na(ald)), statistic = "F")
F = 1.6498, ndf = 2.751, ddf = 112.775, p-value = 0.1858
```

Warning message:
In onestrat(x[index, , drop = FALSE], clusters[index], nPSU[index][1], 
Stratum (18) has only one PSU at stage 1

# ANALYSIS EXAMPLE 6.9 LOGISTIC REGRESSION OF MDE ON MALE NCS-R PART 1 WEIGHT

```r
> (ex69 <- svyglm (mde~sexm, design=ncsrsvyp1, family=quasibinomial))
```

```r
Stratified 1 - level Cluster Sampling design (with replacement)
With (84) clusters.
```

```r
svydesign(strata = ~SESTRAT, id = ~SECLUSTR, weights = ~NCSRWTSH, 
data = ncsr, nest = T)
```

```r
Call:  svyglm(mde ~ sexm, design = ncsrsvyp1, family = quasibinomial)
```

```r
Coefficients:
 (Intercept)         sexm
 -1.21222    0.51617
```

```r
Degrees of Freedom:  9281 Total (i.e. Null);  41 Residual
Null Deviance:  0.9774
Residual Deviance:  0.9674     AIC: NA
```

```r
> summary(ex69)
```

```r
Call:  svyglm(mde ~ sexm, design = ncsrsvyp1, family = quasibinomial)
```

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

```r
Coefficients:
 Estimate Std. Error t value Pr(>|t|)
(Intercept) -1.21222  0.03195  -37.935  < 2e-16 ***
sexm         -0.51617  0.06820  -7.568 2.63e-09 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

```r
(Dispersion parameter for quasibinomial family taken to be 1.000108)
```

```r
Number of Fisher Scoring iterations: 4
```
# FIGURE 6.8 BLOOD PRESSURE STATUS BY GENDER NHANES ADULT DATA
# CREATE OBJECT TO USE IN BARPLOT

fig68 <- svyby(~factor(bp_cat) ~RIAGENDR, subnhanes, svymean, na.rm=T)

> print(fig68)

          RIAGENDR factor(bp_cat)1 factor(bp_cat)2 factor(bp_cat)3 factor(bp_cat)4 se.factor(bp_cat)1 se.factor(bp_cat)2 se.factor(bp_cat)3 se.factor(bp_cat)4
1        1       0.4003467       0.4985263      0.08376503    0.01736189           0.01392390      0.01419122       0.009981295 0.003365517
2        2       0.5372640       0.3437020      0.08888301    0.03015093           0.01325454      0.01430770       0.006790864 0.002765768

> barplot(fig68, legend=c("Normal", "Pre-HBP", "Stage 1 HBP", "Stage 2 HBP"), col=c("black", "grey60", "blue", "red"), xlab=c("Male","Female"))