

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 5 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, svytotal, svyratio and svycontrast are all used to analyze continuous variables/data. In addition, use of svyhist, svyboxplot, and svyplot commands for graphics are demonstrated. Other features included in this chapter are use of a factor statement for categorical variables, the confint command for confidence intervals, and the SE for extracting statistics from an object.

Prior to running the analysis examples, set up of the data sets and complex sample design objects is demonstrated. This syntax is repeated in each chapter.

```

#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 )
names(nhanesdata)

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

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

#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
#both hh and r weights are needed plus financial respondent for hh level analysis
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)
hrssvsub <-subset(hrssvyhh, KFINR==1)

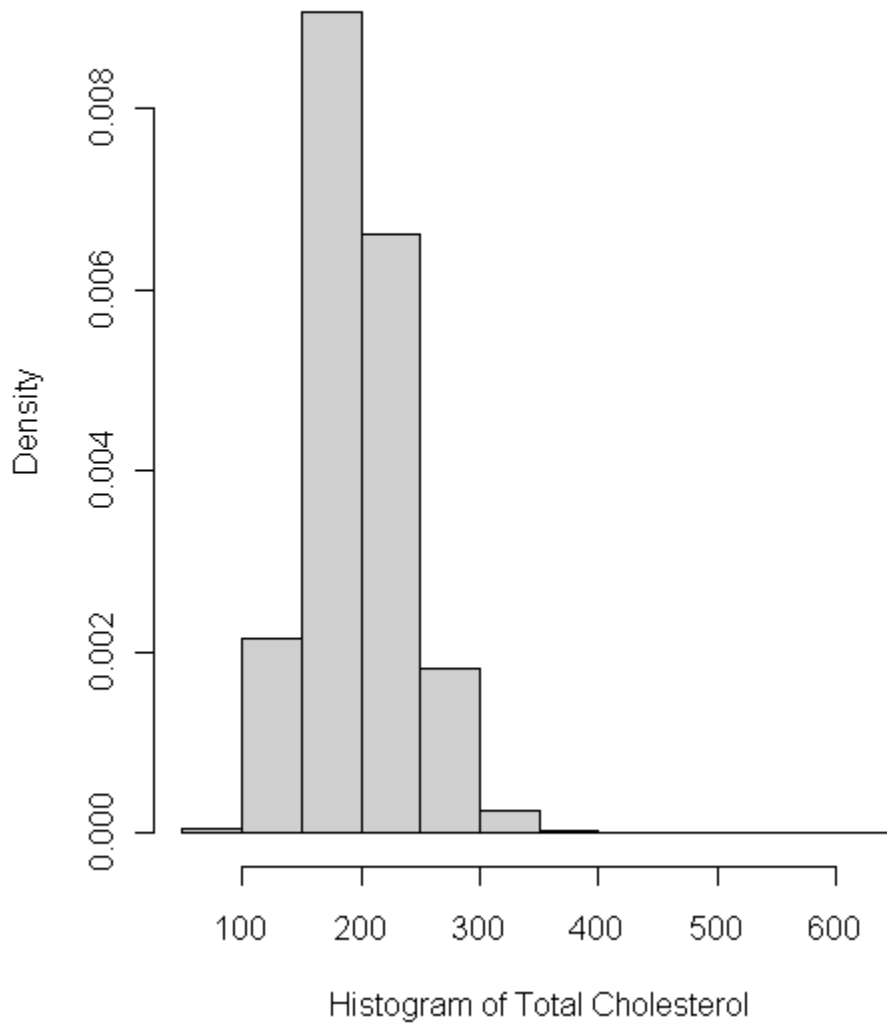
hrssvyr <- svydesign(strata=~STRATUM, id=~SECU, weights=~KWGTR , data=hrs, nest=T)
summary(hrssvyr)

```

```
# CHAPTER 5 ANALYSIS EXAMPLES IN R
```

```
# EXAMPLE 5.1 AND 5.2 ARE GRAPHS USING NHANES ADULT DATA
```

```
svyhist(~LBXTC , subset (nhanessvy2, RIDAGEYR >=18), main="", col="grey80", xlab ="Histogram of Total  
Cholesterol")
```



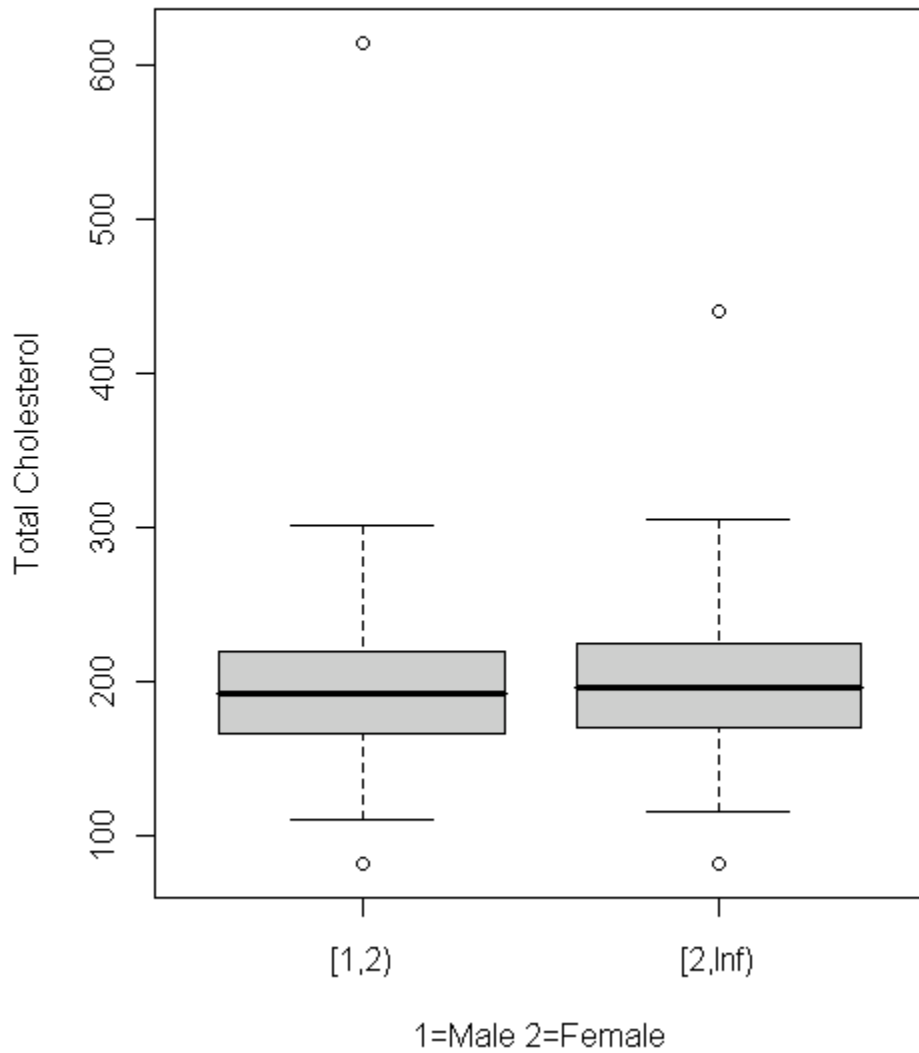
```
#EXAMPLE 5.2 BOXPLOT OF TOTAL CHOLESTEROL BY GENDER NHANES DATA
```

```
#CREATE A VARIABLE CALLED GENDER FOR BOXPLOT
```

```
nhanessvy2<-update(nhanessvy2, gender=cut(RIAGENDR, c(1, 2, Inf), right=F))
```

```
#USE NEW VARIABLE IN BOX PLOT GRAPH
```

```
svyboxplot(LBXTC~gender , subset (nhanessvy2, RIDAGEYR >=18), col="grey80", ylab="Total Cholesterol", xlab  
="1=Male 2=Female")
```



```

# EXAMPLE 5.3 MDE USING POPULATION WEIGHT NCS-R DATA
> svytotal (~mde, ncsrsvypop, deff=T)
      total      SE  DEff
mde 40092207 2567488 9.028

> confint(svytotal(~mde, ncsrsvypop))
      2.5 %  97.5 %
mde 35060023 45124391

#MDE OVER MARITAL STATUS
> ex53 <- svyby (~mde, ~mar3catc, ncsrsvypop, svytotal, deff=T)

> ex53
      mar3catc      mde      se.mde  DEff.mde
Married              Married 20304191 1584108.7 6.817920
Previously Married Previously Married 10360671 702621.5 2.966192
Never Married              Never Married 9427346 773137.6 3.063915

> confint(ex53)
      2.5 %  97.5 %
Married      17199395 23408987
Previously Married 8983558 11737784
Never Married    7912024 10942667
Warning message:
In vcov.svyby(object) : Only diagonal elements of vcov() available

#Example 5.4 HRS DATA
> svyby (~H8ATOTA, ~I(KFINR==1), hrssvy, na.rm=T, svytotal)
      I(KFINR == 1)      H8ATOTA      se.H8ATOTA
FALSE      FALSE 2.027941e+13 1.323100e+12
TRUE       TRUE 2.839749e+13 1.595586e+12

> confint(svyby (~H8ATOTA, ~I(KFINR==1), hrssvy, na.rm=T, ci=T, svytotal))
      2.5 %      97.5 %
FALSE 1.768618e+13 2.287263e+13
TRUE 2.527020e+13 3.152478e+13
Warning message:
In vcov.svyby(object) : Only diagonal elements of vcov() available

```

```
#EXAMPLE 5.5 NCS-R DATA WITH PART 2 WEIGHT
```

```
> svymean(~HHINC, design=ncsrsvyp2wt, na.rm=T, deff="replace")
      mean      SE  DEff
HHINC 59277.1 1596.3 6.0938
```

```
#CREATE OBJECT CALLED OB AND THEN EXTRACT CONFIDENCE INTERVALS
```

```
> ob <- svymean(~HHINC, ncsrsvyp2wt, na.rm=T, deff="replace")
```

```
> confint(ob)
      2.5 %  97.5 %
HHINC 56148.28 62405.83
```

```
#EXAMPLE 5.6 NHANES DATA
```

```
> a <- svymean(~BPXSY1, subset(nhanessvy2, RIDAGEYR >=18), na.rm=TRUE)
```

```
> coef(a)
BPXSY1
123.1109
```

```
> SE(a)
BPXSY1
0.5416936
```

```
> confint(a)
      2.5 %  97.5 %
BPXSY1 122.0492 124.1726
```

```
#EXAMPLE 5.7 HRS DATA, USES INDICATOR VARIABLE TO REPRESENT FINANCIAL RESPONDENT TRUE/FALSE
```

```
> svyby(~H8ATOTA, ~I(KFINR==1), hrssvy, na.rm=T, ci=T, svymean)
```

```
      I(KFINR == 1)  H8ATOTA  se.H8ATOTA
FALSE      FALSE 714161.2  44891.87
TRUE       TRUE  527313.2  28012.78
```

```
> confint(svyby(~H8ATOTA, ~I(KFINR==1), hrssvy, na.rm=T, ci=T, svymean))
```

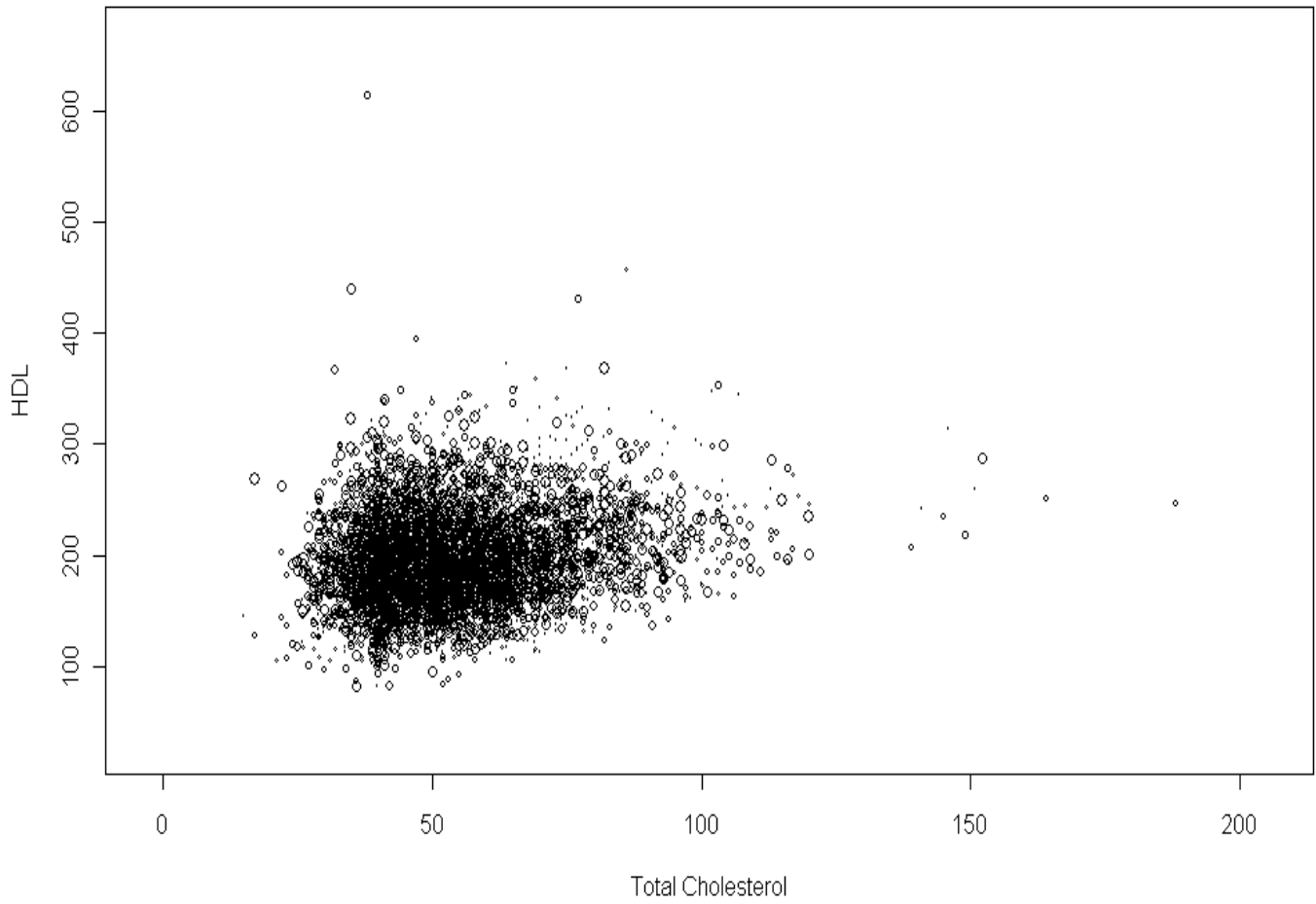
```
      2.5 %  97.5 %
FALSE 626174.7 802147.6
TRUE  472409.1 582217.2
```

```
Warning message:
```

```
In vcov.svyby(object) : Only diagonal elements of vcov() available
```

```
#Scatter Plot of HDL and Total Cholesterol NHANES data
```

```
svyplot(LBXTCLBDHDD, subset(nhanessvy2, RIDAGEYR >=18), style="bubble", ylab="HDL", xlab="Total Cholesterol")
```



```
#EXAMPLE 5.9 NHANES DATA WITH INDICATOR OF ADULT (TRUE/FALSE)
```

```
> svyby (~LBDHDD, denominator=~LBXTC, by=~I(RIDAGEYR >= 18), nhanessvy2, na.rm=T, ci=T, svyratio)
```

	I(RIDAGEYR >= 18)	V1	se
FALSE	FALSE	0.3355439	0.002475870
TRUE	TRUE	0.2753546	0.002254350

```
> confint(svyby (~LBDHDD, denominator=~LBXTC, by=~I(RIDAGEYR >= 18), nhanessvy2, na.rm=T, ci=T, svyratio))
```

	2.5 %	97.5 %
FALSE	0.3306913	0.3403965
TRUE	0.2709362	0.2797730

Warning message:

In vcov.svyby(object) : Only diagonal elements of vcov() available

```
#EXAMPLE 5.10 HRS DATA DIABETES BY GENDER WITH SUBSET OF THOSE AGED OVER 70 YEARS
```

```
> hrssvyr <- svydesign(strata=~STRATUM, id=~SECU, weights=~KWGTR, data=hrs, nest=T, na.rm=T)
```

```
#USE OF SUBSET (PROVIDES CORRECT SUBSET OF DATA WHILE STILL USING ENTIRE SAMPLE COMPLEX DESIGN FEATURES, SIMILAR TO SUBPOP OR DOMAIN STATEMENT OF STATA AND SAS)
```

```
> subhrs <- subset(hrssvyr, KAGE > 70)
```

```
> by <- svyby(~DIABETES, ~GENDER, subhrs, svymean, keep.names=T, na.rm=T)
```

```
#CODES FOR GENDER 1=MALE 2=FEMALE
```

```
> print(by, digits=3)
```

	GENDER	DIABETES	se.DIABETES
1	1	0.235	0.00832
2	2	0.184	0.00853

```
> confint(by)
```

	2.5 %	97.5 %
1	0.2190736	0.2516751
2	0.1671450	0.2005859

Warning message:

In vcov.svyby(object) : Only diagonal elements of vcov() available

```
#EXAMPLE 5.11 NHANES DATA BLOOD PRESSURE BY GENDER IN SUBSET OF THOSE >= 46 YEARS OF AGE
```

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

```
> subnhanes46 <- subset(nhanessvy2, RIDAGEYR >= 46)
```

```
#CODES FOR RIAGENDR 1=MALE 2=FEMALE
```

```
> (ex511 <- svyby(~BPXSY1, ~RIAGENDR, subnhanes46, svymean, keep.names=T, na.rm=T))
```

	RIAGENDR	BPXSY1	se.BPXSY1
1	1	128.9629	0.7566666
2	2	132.0873	1.0645383

```
> confint(ex511)
```

	2.5 %	97.5 %
1	127.4799	130.4460
2	130.0008	134.1737

Warning message:

In vcov.svyby(object) : Only diagonal elements of vcov() available

#EXAMPLE 5.12 HRS DATA TOTAL HOUSEHOLD ASSETS BY EDUCATION

#CODES FOR EDCAT: 1=0-11 2=12 3=13-15 4=16+ YEARS OF EDUCATION

```
> ex512 <- svyby(~H8ATOTA, ~EDCAT, subhrsfinr, svymean, na.rm=T, options(survey.lonely.psu="remove"))
```

There were 22 warnings (use warnings() to see them)

#CODES FOR EDCAT 1=0-11 2=12 3=13-15 4=16+ (COULD ALSO USE THE FACTOR VARIABLE APPROACH)

```
> print(ex512)
```

	EDCAT	H8ATOTA	se.H8ATOTA
1	1	178386.1	24561.12
2	2	328392.0	17082.72
3	3	455457.6	27000.33
4	4	1107204.1	102113.52
NA	NA	153056.4	35184.50

```
> confint(ex512)
```

	2.5 %	97.5 %
1	130247.14	226525.0
2	294910.45	361873.5
3	402537.93	508377.3
4	907065.28	1307342.9
NA	84096.06	222016.8

Warning message:

In vcov.svyby(object) : Only diagonal elements of vcov() available

```
> svycontrast(ex512, list(avg=c(.5,0,0,.5,0), diff=c(1,0,0,-1,0)))
```

	contrast	SE
avg	642795	52513
diff	-928818	105026

Warning message:

In vcov.svyby(stat) : Only diagonal elements of vcov() available

#EXAMPLE 5.13 HRS 2004 AND 2006 DATA TOTAL HH ASSETS BY YEAR

```
> ex513 <- svyby(~totassets, ~year, design=sub0406finr, keep.vars=T, svymean)
```

```
> coef(ex513)
```

	2004	2006
	411786.4	527313.2

```
> SE(ex513)
```

[1] 20639.72 28012.78

```
> svycontrast(ex513, list(avg=c(.5,.5), diff=c(1,-1)))
```

	contrast	SE
avg	469550	17398
diff	-115527	34795

Warning message:

In vcov.svyby(stat) : Only diagonal elements of vcov() available