CHAPTER 10 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.

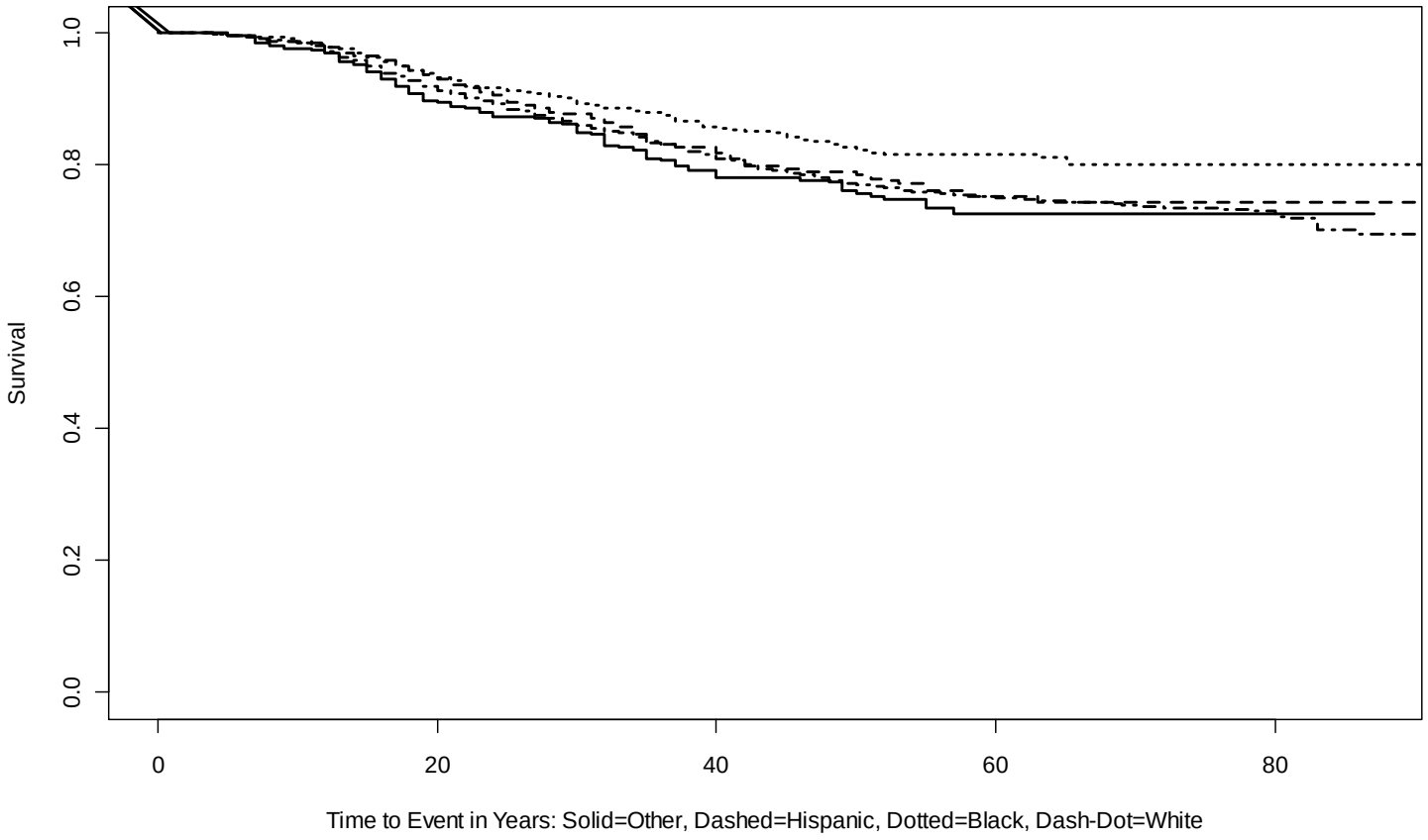
NOTES ABOUT SURVIVAL ANALYSES IN R SURVEY PACKAGE 3.22 (WITH R 2.7)

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 full range of svy commands for survival analysis: svykm for survival curves along with the plot command to produce high quality plots of the object created by svykm, and the svycoxph command for Cox Proportional Hazards models are demonstrated. For discrete-time logistic and Clog-log regression, use of svyglm with how to read in a person year data set is shown. Use of the correct family and link options is specified within svyglm in order to obtain the correct model.

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

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

```
#EXAMPLE OF SURVIVAL CURVE USING SVYKM

(kmsvy <- svykm(Surv(ageonsetmde,mde)~strata(racecat), design=ncsrsvyp1))

plot(kmsvy,lwd=2,pars=list(lty=c(1,2,3,4)),ylab=c("Survival"),xlab=c("Time to Event in Years: Solid=Other,
Dashed=Hispanic, Dotted=Black, Dash-Dot=White"))

# VARIABLE PLOTTED IS RACECAT CODES 1=OTHER/ASIAN 2=HISPANIC 3=BLACK 4=WHITE
```



```
#COX PH MODEL NCS-R DATA

ex102_coxph<-svycoxph(Surv(ageonsetmde,mde)~AGE + sexm +
factor(MAR3CAT)+factor(ED4CAT)+factor(racecat),design=ncsrsvyp1)

> summary(ex102_coxph)

Stratified 1 - level Cluster Sampling design (with replacement)
With (84) clusters.
svydesign(strata = ~SESTRAT, id = ~SECLUSTR, weights = ~NCSRWTSH,
    data = ncsr, nest = T)
Call:
svycoxph.survey.design(formula = Surv(ageonsetmde, mde) ~ AGE +
    sexm + factor(MAR3CAT) + factor(ED4CAT) + factor(racecat),
    design = ncsrsvyp1)  n= 9282


#NOTE CODES FOR FACTOR VARIABLES
MAR3CAT 1=MARRIED 2=PREVIOUSLY MARRIED 3=NEVER MARRIED
ED4CAT 1=0-11 2=12 3=13-15 4=16+ YEARS OF EDUCATION
RACECAT 1=OTHER 2=HISPANIC 3=BLACK 4=WHITE

                     coef exp(coef) se(coef)       z       p
AGE              -0.0497     0.952  0.00239 -20.766 0.0e+00
sexm             -0.4554     0.634  0.06254  -7.281 3.3e-13
factor(MAR3CAT)2  0.5047     1.657  0.06034   8.364 1.1e-16
factor(MAR3CAT)3  0.0815     1.085  0.08918   0.914 3.6e-01
factor(ED4CAT)2  -0.0574     0.944  0.06736  -0.853 3.9e-01
factor(ED4CAT)3   0.0451     1.046  0.05831   0.774 4.4e-01
factor(ED4CAT)4  -0.0915     0.913  0.06393  -1.430 1.5e-01
factor(racecat)2 -0.2514     0.778  0.13517  -1.860 6.3e-02
factor(racecat)3 -0.4811     0.618  0.14979  -3.212 1.3e-03
factor(racecat)4  0.0782     1.081  0.11822   0.661 5.1e-01

                 exp(coef) exp(-coef) lower .95 upper .95
AGE                  0.952      1.051     0.947     0.956
sexm                 0.634      1.577     0.561     0.717
factor(MAR3CAT)2     1.657      0.604     1.472     1.864
factor(MAR3CAT)3     1.085      0.922     0.911     1.292
factor(ED4CAT)2      0.944      1.059     0.827     1.077
factor(ED4CAT)3      1.046      0.956     0.933     1.173
factor(ED4CAT)4      0.913      1.096     0.805     1.034
factor(racecat)2     0.778      1.286     0.597     1.014
factor(racecat)3     0.618      1.618     0.461     0.829
factor(racecat)4     1.081      0.925     0.858     1.363

Rsquare= NA   (max possible= NA )
Likelihood ratio test= NA  on 10 df,   p=NA
Wald test            = 672  on 10 df,   p=0
Score (logrank) test = NA  on 10 df,   p=NA


#TEST OF RACE ADJUSTED BY OTHER COVARIATES NOT INCLUDED IN THIS PROCEDURE
```

```
#DISCRETE TIME LOGISTIC REGRESSION NCS-R DATA
#EXAMPLE 10.5 LOGIT

> #discrete time logistic using NCS-R data in person year format
> #read in personyear data created in external package (SAS)

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

# survey design
> ncsrsvypyp1 <- svydesign(strata=~SESTRAT, id=~SECLUSTR, weights=~NCSRWTSH, data=ncsrpy, nest=T)

# variable names
> names(ncsrpy)
 [1] "CASEID"      "SC9_1"       "DSM_AGO"     "DSM_ALA"     "DSM_GAD"
 [6] "DSM_PDS"     "DSM_PTS"     "DSM_SO"      "DSM_SP"      "GAD_OND"
[11] "MDE_OND"     "AGE"         "REGION"      "POVINDEX"    "MAR3CAT"
[16] "ED4CAT"      "OBESE6CA"    "NEWSOCIA"    "NCSRWTSH"    "NCSRWTLG"
[21] "SEX"         "HHINC"       "WKSTAT3C"    "WEIGHT"      "HEIGHT"
[26] "RANCEST"     "SESTRAT"     "SECLUSTR"    "ageonsetmde" "bmi"
[31] "mde"         "anyanx"      "sexf"        "sexm"        "ald"
[36] "racecat"     "povcat"      "agecentered" "int"         "mdetv"
[41] "gadtv"       "ageonsetgad" "intwage"

#subset of person years up to and including the year of event or censor (Time at Risk)
> subncsrpy <- subset(ncsrsvypyp1, int <= ageonsetmde)


# logit model
> summary(ex105_logit <- svyglm(mdetv ~ int + AGE + sexm + factor(ED4CAT) + factor(racecat) + factor(MAR3CAT),
family=quasibinomial, design=subncsrpy))
Call:
svyglm(mdetv ~ int + AGE + sexm + factor(ED4CAT) + factor(racecat) +
    factor(MAR3CAT), family = quasibinomial, design = subncsrpy)

Survey design:
subset(ncsrsvypyp1, int <= ageonsetmde)

#NOTE CODES FOR FACTOR VARIABLES
MAR3CAT 1=MARRIED 2=PREVIOUSLY MARRIED 3=NEVER MARRIED
ED4CAT 1=0-11 2=12 3=13-15 4=16+ YEARS OF EDUCATION
RACECAT 1=OTHER 2=HISPANIC 3=BLACK 4=WHITE

Coefficients:
                  Estimate Std. Error t value Pr(>|t|)
(Intercept)      -3.435525   0.161987 -21.209  < 2e-16 ***
int               0.032798   0.002074  15.816  < 2e-16 ***
AGE              -0.058334   0.002449 -23.823  < 2e-16 ***
sexm             -0.444869   0.062288  -7.142 5.00e-08 ***
factor(ED4CAT)2  -0.020136   0.066114  -0.305  0.76273
factor(ED4CAT)3   0.092919   0.057444   1.618  0.11589
factor(ED4CAT)4  -0.019451   0.063338  -0.307  0.76082
factor(racecat)2 -0.248422   0.134769  -1.843  0.07486 .
factor(racecat)3 -0.456968   0.149888  -3.049  0.00467 **
factor(racecat)4  0.073996   0.118239   0.626  0.53602
factor(MAR3CAT)2  0.494250   0.061010   8.101 3.78e-09 ***
factor(MAR3CAT)3 -0.035346   0.087970  -0.402  0.69059
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for quasibinomial family taken to be 1.002008)

Number of Fisher Scoring iterations: 9
There were 12 warnings (use warnings() to see them)
```

```
# obtain odds ratios

> exp(ex105_logit$coef)
     (Intercept)             int             AGE            sexm
      0.03220851       1.03334155       0.94333508       0.64090809
 factor(ED4CAT)2  factor(ED4CAT)3  factor(ED4CAT)4 factor(racecat)2
      0.98006512       1.09737261       0.98073699       0.78003095
factor(racecat)3 factor(racecat)4 factor(MAR3CAT)2 factor(MAR3CAT)3
      0.63320074       1.07680197       1.63926854       0.96527120
```

```
#CLOGLOG EXAMPLE 10.5

> (ex105_cloglog<-svyglm(mdetv ~ int + AGE + sexm + factor(ED4CAT) + factor(racecat) + factor(MAR3CAT),
family=quasibinomial (link=cloglog), design=subncsrpy))
Stratified 1 - level Cluster Sampling design (with replacement)
With (84) clusters.
subset(ncsrsvypyp1, int <= ageonsetmde)

Call:  svyglm(mdetv ~ int + AGE + sexm + factor(ED4CAT) + factor(racecat) +      factor(MAR3CAT), family =
quasibinomial(link = cloglog),      design = subncsrpy)

Coefficients:
     (Intercept)              int              AGE             sexm
        -3.44439          0.03273         -0.05818         -0.44322
 factor(ED4CAT)2  factor(ED4CAT)3  factor(ED4CAT)4 factor(racecat)2
        -0.01974          0.09236         -0.01920         -0.24742
factor(racecat)3 factor(racecat)4 factor(MAR3CAT)2 factor(MAR3CAT)3
        -0.45508          0.07373          0.49281         -0.03547

Degrees of Freedom: 385695 Total (i.e. Null);  31 Residual
Null Deviance:      0.05869
Residual Deviance: 0.05598      AIC: NA

> summary(ex105_cloglog)
Call:
svyglm(mdetv ~ int + AGE + sexm + factor(ED4CAT) + factor(racecat) +
    factor(MAR3CAT), family = quasibinomial(link = cloglog),
    design = subncsrpy)

Survey design:
subset(ncsrsvypyp1, int <= ageonsetmde)

#NOTE CODES FOR FACTOR VARIABLES
MAR3CAT 1=MARRIED 2=PREVIOUSLY MARRIED 3=NEVER MARRIED
ED4CAT 1=0-11 2=12 3=13-15 4=16+ YEARS OF EDUCATION
RACECAT 1=OTHER 2=HISPANIC 3=BLACK 4=WHITE

Coefficients:
                  Estimate Std. Error t value Pr(>|t|)
(Intercept)      -3.444394   0.161368 -21.345  < 2e-16 ***
int               0.032733   0.002069  15.822  < 2e-16 ***
AGE              -0.058180   0.002440 -23.842  < 2e-16 ***
sexm             -0.443221   0.062082  -7.139 5.04e-08 ***
factor(ED4CAT)2  -0.019740   0.065853  -0.300  0.76636
factor(ED4CAT)3   0.092360   0.057197   1.615  0.11649
factor(ED4CAT)4  -0.019204   0.063098  -0.304  0.76290
factor(racecat)2 -0.247424   0.134362  -1.841  0.07514 .
factor(racecat)3 -0.455078   0.149438  -3.045  0.00471 **
factor(racecat)4  0.073734   0.117876   0.626  0.53620
factor(MAR3CAT)2  0.492815   0.060769   8.110 3.70e-09 ***
factor(MAR3CAT)3 -0.035473   0.087538  -0.405  0.68809
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for quasibinomial family taken to be 1.001771)
Number of Fisher Scoring iterations: 9

> exp(exp105_cloglog$coef)
Error: object "exp105_cloglog" not found
> exp(ex105_cloglog$coef)
     (Intercept)              int              AGE             sexm
      0.03192411       1.03327450       0.94347978       0.64196512
 factor(ED4CAT)2  factor(ED4CAT)3  factor(ED4CAT)4 factor(racecat)2
      0.98045358       1.09675913       0.98097958       0.78080930
factor(racecat)3 factor(racecat)4 factor(MAR3CAT)2 factor(MAR3CAT)3
      0.63439833       1.07652090       1.63691738       0.96514832
```