R Analysis Example Replication C10

```
# ASDA2 Chapter 10 Survival Analysis
library(survey)
# Read in C10 data set, this data is set up for survival analysis in one record per person format
ncsrc10 <- read.table(file = "P:/ASDA 2/Data sets/ncsr/c10_ncsr.csv", sep = ",", header = T, as.is=T)
names(ncsrc10)
#create factor versions with labels
ncsrc10$racec <- factor(ncsrc10$racecat, levels = 1: 4, labels =c("Other", "Hispanic", "Black", "White"))
ncsrc10$mar3catc <- factor(ncsrc10$MAR3CAT, levels = 1: 3, labels =c("Married", "Previously Married", "Never
Married"))
ncsrc10$ed4catc <- factor(ncsrc10$ED4CAT, levels = 1: 4, labels =c("0-11", "12", "13-15","16+"))
ncsrc10$sexc <- factor(ncsrc10$SEX, levels = 1:2, labels=c("Male","Female"))
ncsrc10$ag4catc <- factor(ncsrc10$ag4cat, levels = 1:4, labels=c("18-29", "30-44", "45-59", "60+"))
ncsrc10$mdec <- factor(ncsrc10$mde, level = 1:2, labels=c("No","Yes"))

# survey design for one record per person
ncsrsvyc10 <- svydesign(strata=~SESTRAT, id=~SECLUSTR, weights=~NCSRWTSH, data=ncsrc10, nest=T)
names (ncsrsvyc10)
# Example 10.3 KM curve NCSR data, note use of survfit since we do not need SE's for this analysis
(km <- survfit(Surv(ageonsetmde,mde)~strata(racecat), data=ncsrc10, weight=NCSRWTSH))
plot(km,lwd=5,lty=c(1,2,3,4),col=c("blue","green","red", "purple"), ylab=c("Survival"), xlab=c("Time to Event in
Years:  Blue:Other   Green:Hispanic   Red:AfAm   Purple:White"))

# svykm instead for comparison and example
# Note that when using "se=T" it causes R program to stall and die, omit here as PC runs out of memory, see
documentation for details on this issue
(kmsvy <- svykm(Surv(ageonsetmde,mde)~strata(racecat),design=ncsrsvyc10))
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"))

# Example 10.4 Cox model
summary(ex104_coxph<-svycoxph(Surv(ageonsetmde,mde)~intwage + sexm + mar3catc + ed4catc +
racec,design=ncsrsvyc10))
# No test of proportional hazards for race in R
#discrete time logistic using ncsr data in person year format
#read in personyear data, previously set up with multiple records per person
ncsrpy <- read.table(file = "P:/ASDA 2/Data sets/ncsr/c10_expanded1.csv", sep = ",", header = T, as.is=T)
names(ncsrpy)
ncsrsvypyp1 <- svydesign(strata=~SESTRAT, id=~SECLUSTR, weights=~NCSRWTSH, data=ncsrpy, nest=T)

# Example 10.5 discrete time logistic
# Subset of records <= age of onset of mde/censor, needed for model to follow
subncsrpy <- subset(ncsrsvypyp1, pyr <= ageonsetmde)
summary(ex105_logit <- svyglm(mdetv ~ pyr + intwage + sexm + factor(ED4CAT) + factor(racecat) + factor(MAR3CAT),
family=quasibinomial, design=subncsrpy))
# get exponents of betas
exp(ex105_logit$coef)
# With cloglog link
summary(ex105_cloglog<-svyglm(mdetv ~ pyr + intwage + sexm + factor(ED4CAT) + factor(racecat) + factor(MAR3CAT),
family=quasibinomial(link=cloglog), design=subncsrpy))
# With exponentiated coefficients
exp(ex105_logit$coef)
```
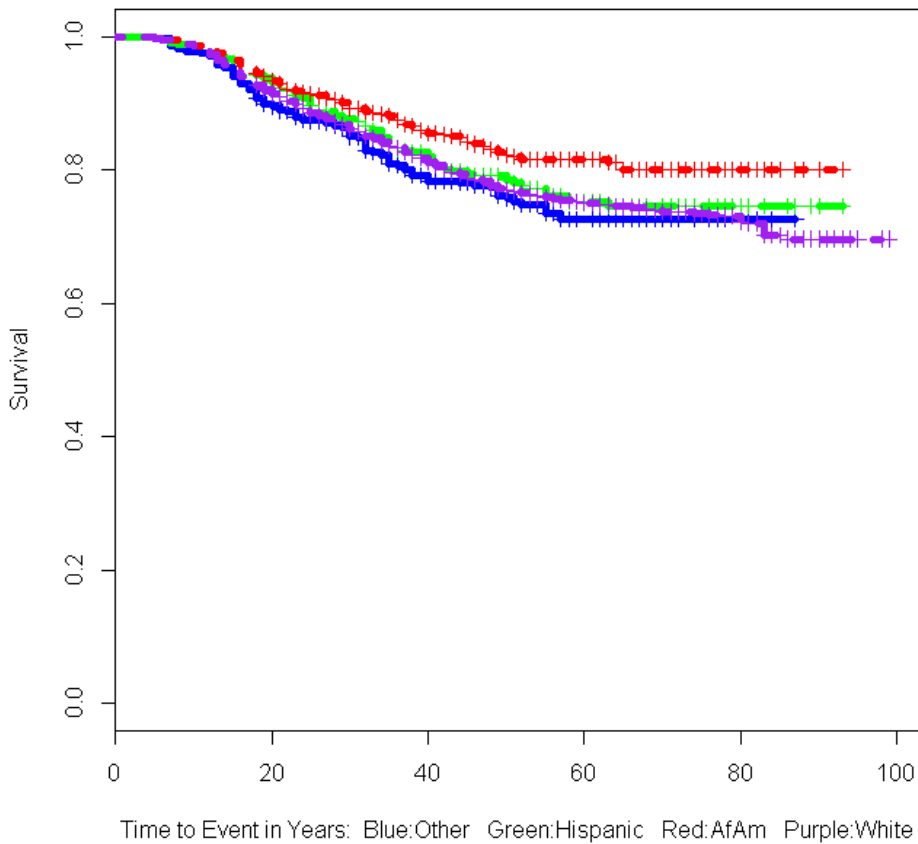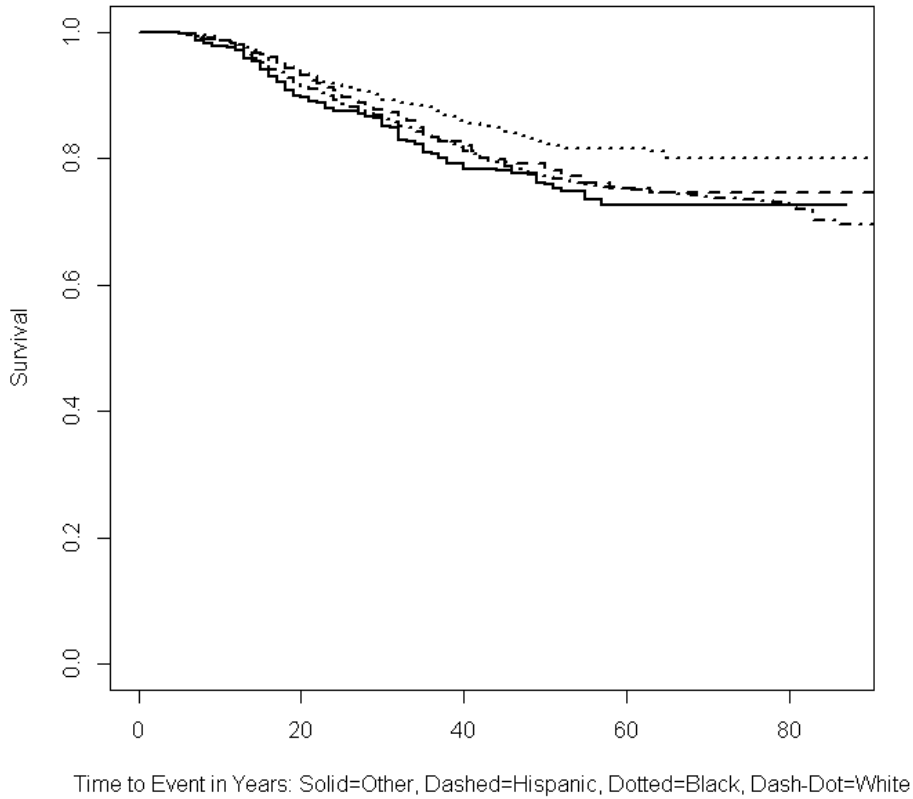
Output R Analysis Example Replication C10

```
> # KM curve NCSR data, note use of survfit since we do not need SE's for this analysis
> (km <- survfit(Surv(ageonsetmde,mde)~strata(racecat), data=ncsrc10, weight=NCSRWTSH))
Call: survfit(formula = Surv(ageonsetmde, mde) ~ strata(racecat), data = ncsrc10,
    weights = NCSRWTSH)
                            records n.max n.start events median 0.95LCL 0.95UCL
strata(racecat)=racecat=1       473   404     404   81.7     NA      NA      NA
strata(racecat)=racecat=2       883  1007    1007  164.9     NA      NA      NA
strata(racecat)=racecat=3      1230  1073    1073  151.0     NA      NA      NA
strata(racecat)=racecat=4      6696  6798    6798 1381.9     NA      NA      NA
> plot(km,lwd=5,lty=c(1,2,3,4),col=c("blue","green","red", "purple"), ylab=c("Survival"), xlab=c("Time to Event
in Years:  Blue:Other   Green:Hispanic   Red:AfAm   Purple:White"))
```


Time to Event in Years: Blue:Other   Green:Hispanic   Red:AfAm   Purple:White

```
#use of svykm instead for comparison and example

(kmsvy <- svykm(Surv(ageonsetmde,mde)~strata(racecat), design=ncsrsvyc10))

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"))
```


Time to Event in Years: Solid=Other, Dashed=Hispanic, Dotted=Black, Dash-Dot=White

```
> # Example 10.4 Cox model
> summary(ex104_coxph<-svycoxph(Surv(ageonsetmde,mde)~intwage + sexm + mar3catc + ed4catc +
racec,design=ncsrsvyc10))
Stratified 1 - level Cluster Sampling design (with replacement)
With (84) clusters.
svydesign(strata = ~SESTRAT, id = ~SECLUSTR, weights = ~NCSRWTSH,
    data = ncsrc10, nest = T)
Call:
svycoxph(formula = Surv(ageonsetmde, mde) ~ intwage + sexm +
    mar3catc + ed4catc + racec, design = ncsrsvyc10)

  n= 9282, number of events= 1829


                              coef exp(coef)
intwage                   -0.049680  0.951534
sexm                      -0.455350  0.634226
mar3catcPreviously Married 0.504709  1.656503
mar3catcNever Married      0.081532  1.084948
ed4catc12                 -0.057437  0.944181
ed4catc13-15               0.045108  1.046141
ed4catc16+                -0.091455  0.912603
racecHispanic             -0.251413  0.777701
racecBlack                -0.481060  0.618128
racecWhite                 0.078158  1.081294
                           se(coef)       z Pr(>|z|)
intwage                    0.002392 -20.766  < 2e-16
sexm                       0.062540  -7.281 3.31e-13
mar3catcPreviously Married 0.060340   8.364  < 2e-16
mar3catcNever Married      0.089182   0.914  0.36060
ed4catc12                  0.067355  -0.853  0.39380
ed4catc13-15               0.058314   0.774  0.43921
ed4catc16+                 0.063933  -1.430  0.15258
racecHispanic              0.135175  -1.860  0.06290
racecBlack                 0.149788  -3.212  0.00132
racecWhite                 0.118217   0.661  0.50852

intwage                    ***
sexm                       ***
mar3catcPreviously Married ***
mar3catcNever Married
ed4catc12
ed4catc13-15
ed4catc16+
racecHispanic                .
racecBlack                 **
racecWhite
---
Signif. codes:
0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1


                           exp(coef) exp(-coef)
intwage                       0.9515     1.0509
sexm                          0.6342     1.5767
```

```
mar3catcPreviously Married    1.6565      0.6037
mar3catcNever Married         1.0849      0.9217
ed4catc12                     0.9442      1.0591
ed4catc13-15                  1.0461      0.9559
ed4catc16+                    0.9126      1.0958
racecHispanic                 0.7777      1.2858
racecBlack                    0.6181      1.6178
racecWhite                    1.0813      0.9248
                          lower .95 upper .95
intwage                       0.9471      0.9560
sexm                          0.5611      0.7169
mar3catcPreviously Married    1.4717      1.8645
mar3catcNever Married         0.9110      1.2922
ed4catc12                     0.8274      1.0774
ed4catc13-15                  0.9332      1.1728
ed4catc16+                    0.8051      1.0344
racecHispanic                 0.5967      1.0136
racecBlack                    0.4609      0.8290
racecWhite                    0.8577      1.3632


Concordance= 0.694  (se = 0.007 )
Rsquare= NA    (max possible= NA )
Likelihood ratio test= NA  on 10 df,   p=NA
Wald test            = 672.5  on 10 df,   p=0
Score (logrank) test = NA  on 10 df,   p=NA
```

> # No test of proportional hazards for race in R

```
> #discrete time logistic using NCSR data in person year format
> #read in personyear data, previously set up with multiple records per person
> ncsrpy <- read.table(file = "P:/ASDA 2/Data sets/ncsr/c10_expanded1.csv", sep = ",", header = T, as.is=T)
> names(ncsrpy)
 [1] "CASEID"     "DSM_SO"     "MDE_OND"    "SO_OND"     "AGE"        "REGION"     "MAR3CAT"
 [8] "ED4CAT"     "OBESE6CA"   "NCSRWTSH"   "NCSRWTLG"   "SEX"        "WKSTAT3C"   "SESTRAT"
[15] "SECLUSTR"   "ag4cat"     "racecat"    "mde"        "ald"        "sexf"       "sexm"
[22] "ageonsetmde" "intwage"   "ncsrwtsh100" "pyr"       "mdetv"

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

> # Example 10.5 discrete time logistic
> # Subset of records <= age of onset of mde/censor, needed for model to follow
> subncsrpy <- subset(ncsrsvypyp1, pyr <= ageonsetmde)

> summary(ex105_logit <- svyglm(mdetv ~ pyr + intwage + sexm + factor(ED4CAT) + factor(racecat) +
factor(MAR3CAT), family=quasibinomial, design=subncsrpy))

Call:
svyglm(formula = mdetv ~ pyr + intwage + sexm + factor(ED4CAT) +
    factor(racecat) + factor(MAR3CAT), family = quasibinomial,
    design = subncsrpy)

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

Coefficients:
                Estimate Std. Error t value Pr(>|t|)
(Intercept)     -3.435525   0.161988 -21.209  < 2e-16 ***
pyr              0.032798   0.002074  15.816  < 2e-16 ***
intwage         -0.058334   0.002449 -23.823  < 2e-16 ***
sexm            -0.444869   0.062288  -7.142 5.00e-08 ***
factor(ED4CAT)2 -0.020136   0.066115  -0.305  0.76273
factor(ED4CAT)3  0.092919   0.057445   1.618  0.11589
factor(ED4CAT)4 -0.019451   0.063338  -0.307  0.76082
factor(racecat)2 -0.248422  0.134771  -1.843  0.07487 .
factor(racecat)3 -0.456968  0.149889  -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

> # get exponents of betas
> exp(ex105_logit$coef)
    (Intercept)              pyr          intwage             sexm  factor(ED4CAT)2  factor(ED4CAT)3
     0.03220851       1.03334155       0.94333508       0.64090809       0.98006512       1.09737261
 factor(ED4CAT)4 factor(racecat)2 factor(racecat)3 factor(racecat)4 factor(MAR3CAT)2 factor(MAR3CAT)3
     0.98073699       0.78003095       0.63320074       1.07680197       1.63926854       0.96527120
```

```
 > # With cloglog link
> summary(ex105_cloglog<-svyglm(mdetv ~ pyr + intwage + sexm + factor(ED4CAT) + factor(racecat) +
factor(MAR3CAT), family=quasibinomial(link=cloglog), design=subncsrpy))

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

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

Coefficients:
                Estimate Std. Error t value Pr(>|t|)
(Intercept)     -3.444394   0.161374 -21.344  < 2e-16 ***
pyr              0.032733   0.002069  15.821  < 2e-16 ***
intwage         -0.058180   0.002440 -23.840  < 2e-16 ***
sexm            -0.443221   0.062080  -7.139 5.04e-08 ***
factor(ED4CAT)2 -0.019740   0.065854  -0.300  0.76637
factor(ED4CAT)3  0.092360   0.057200   1.615  0.11651
factor(ED4CAT)4 -0.019204   0.063098  -0.304  0.76290
factor(racecat)2 -0.247424  0.134369  -1.841  0.07515 .
factor(racecat)3 -0.455078  0.149441  -3.045  0.00471 **
factor(racecat)4  0.073735  0.117878   0.626  0.53621
factor(MAR3CAT)2  0.492815  0.060770   8.110 3.70e-09 ***
factor(MAR3CAT)3 -0.035473  0.087535  -0.405  0.68808
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for quasibinomial family taken to be 1.001772)

Number of Fisher Scoring iterations: 9

> # With exponentiated coefficients
> exp(ex105_logit$coef)
    (Intercept)              pyr            intwage             sexm  factor(ED4CAT)2  factor(ED4CAT)3
     0.03220851       1.03334155         0.94333508       0.64090809       0.98006512       1.09737261
 factor(ED4CAT)4 factor(racecat)2 factor(racecat)3 factor(racecat)4 factor(MAR3CAT)2 factor(MAR3CAT)3
     0.98073699       0.78003095         0.63320074       1.07680197       1.63926854       0.96527120
```