

ASDA3 Analysis Examples Replication Chapter 10

R code

```
# Load needed packages
library(survey)
library(haven)
library(sas7bdat)

setwd("P:/ASDA3/Replication R/Chapter 10")

# Read in C10 data set, this SAS data is ready for survival analysis, one record per person format (see text for
details)

ncsrc10 <- read_sas("P:/ASDA3/Replication SAS/c10_ncsr.sas7bdat")
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 wide data
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))

png("P:/ASDA3/Replication R/Chapter 10/fig103_km.png")
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"))
dev.off()

# Use 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"))

# Additional example of use of color along w/line type
(kmsvy <- svykm(Surv(ageonsetmde,mde)~strata(sexf), design=ncsrsvyc10))
plot(kmsvy,lwd=2,pars=list(lty=c(1,2), col=c("black", "red"),ylab=c("Survival"),xlab=c("Time to Event in Years:
Black=M, Red=F"))))

# Example 10.4 Cox model
summary(ex104_coxph<-svycoxph(Surv(ageonsetmde,mde)~ racec + intwage + sexm + mar3catc + ed4catc,
design=ncsrsvyc10))

# No formal test of proportional assumption in R, could try using adjusted model but we have an informal test
from kmsvy above

# Discrete time logistic using ncsr data in person year format
# Read in personyear data, previously set up with multiple records per person (SAS)

ncsrpy <- read_sas("P:/ASDA3/Replication SAS/c10_expanded1.sas7bdat")
names(ncsrpy)
```

```

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

# Example 10.5 DT logistic regression

# Prepare 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 with CIs
coef <- exp(coef(ex105_logit))
coef
ci_exp <- exp(confint(ex105_logit))
ci_exp

# 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
coef <- exp(coef(ex105_cloglog))
coef
ci_exp <- exp(confint(ex105_cloglog))
ci_exp

```

R Results

```
> # Analysis Examples Chapter 10 ASDA3 Survival Analysis
> # Load needed packages
> library(survey)
> library(haven)
> library(sas7bdat)

> # Read in C10 data set, this SAS data is ready for survival analysis, one record per person format
> ncsrc10 <- read_sas("P:/ASDA3/Replication SAS/c10_ncsr.sas7bdat")
> names(ncsrc10)
 [1] "CASEID"      "DSM_SO"      "MDE_OND"     "SO_OND"      "AGE"
 [6] "REGION"     "MAR3CAT"     "ED4CAT"      "OBESE6CA"    "NCSRWTSH"
[11] "NCSRWTLG"   "SEX"         "WKSTAT3C"    "SESTRAT"     "SECLUSTR"
[16] "ag4cat"     "racecat"     "mde"         "ald"         "sexf"
[21] "sexm"       "ageonsetmde" "intwage"     "ncsrwtsh100"

> setwd("P:/ASDA3/Replication R/Chapter 10")

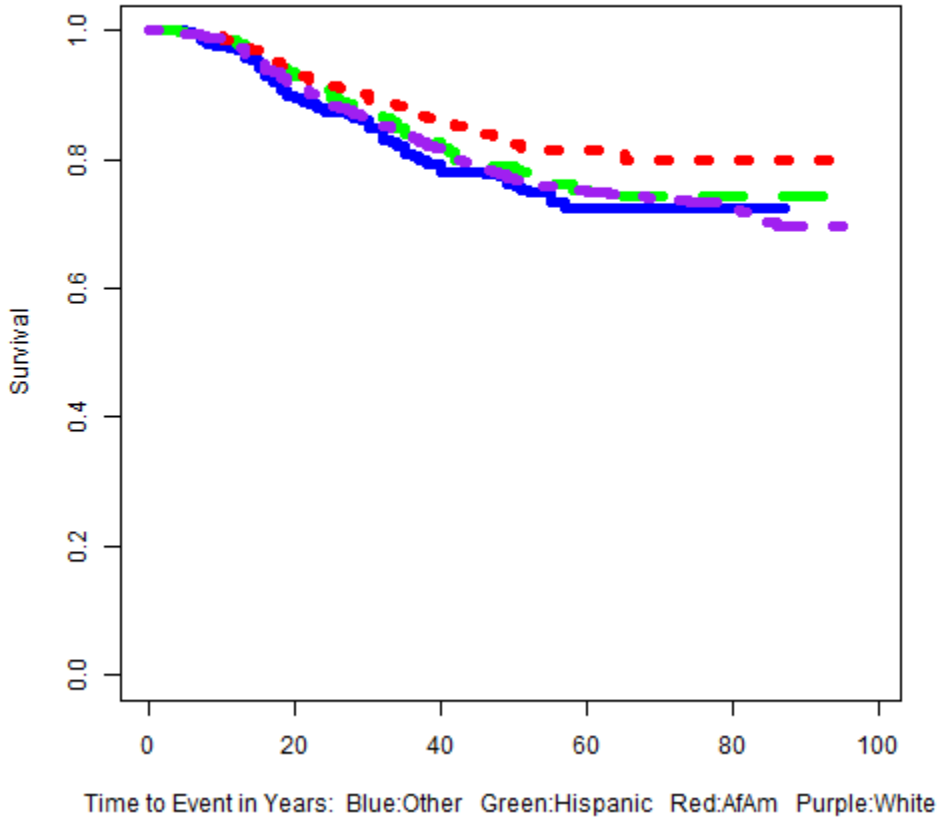
> #create factor versions with labels
> ncsrc10$racec <- factor(ncsrc10$racecat, levels = 1: 4, labels =c("Other", "Hispanic", "Black", "White"))
> ncsrc10$mar3catc <- factor(ncsrc10$MAR3CAT, levels = 1: 3, labels =c("Married", "Previously Married", "Never Married"))
> ncsrc10$ed4catc <- factor(ncsrc10$ED4CAT, 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 wide data
> ncsrsvyc10 <- svydesign(strata=~SESTRAT, id=~SECLUSTR, weights=~NCSRWTSH, data=ncsrc10, nest=T)
> names(ncsrsvyc10)
 [1] "cluster"     "strata"      "has.strata"  "prob"        "allprob"     "call"
 [7] "variables"   "fpc"         "pps"
```

```
> # 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))
Call: survfit(formula = Surv(ageonsetmde, mde) ~ strata(racecat), data = ncsrc10,
               weights = NCSRWTSH)
```

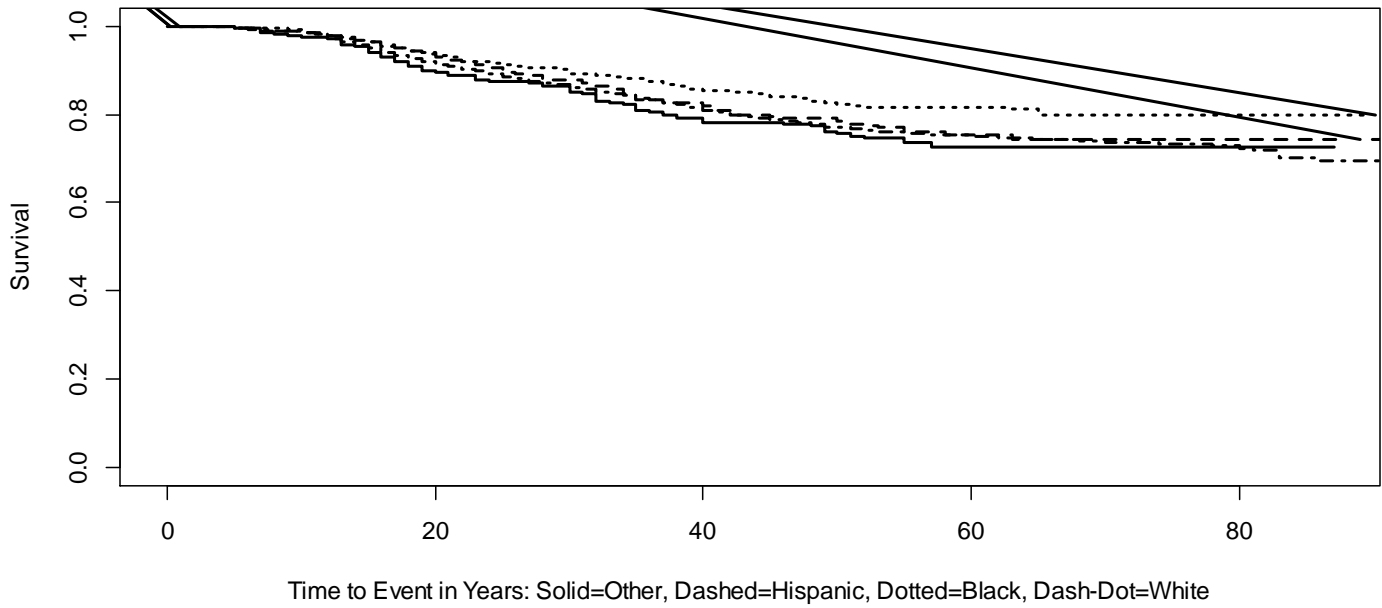
	records	n	events	median	0.95LCL	0.95UCL
strata(racecat)=racecat=1	473	404	81.7	NA	NA	NA
strata(racecat)=racecat=2	883	1007	164.9	NA	NA	NA
strata(racecat)=racecat=3	1230	1073	151.0	NA	NA	NA
strata(racecat)=racecat=4	6696	6798	1381.9	NA	NA	NA

```
> 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"))
```



```
> (kmsvy <- svykm(Surv(ageonsetmde,mde)~strata(racecat),design=ncsrsvyc10))
Weighted survival curves:
svykm(formula = Surv(ageonsetmde, mde) ~ strata(racecat), design = ncsrsvyc10)
racecat=1 : Q1 = 52  median = Inf  Q3 = Inf
racecat=2 : Q1 = 63  median = Inf  Q3 = Inf
racecat=3 : Q1 = Inf  median = Inf  Q3 = Inf
racecat=4 : Q1 = 61  median = Inf  Q3 = Inf

> 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)~ racec + intwage + sexm + mar3catc + ed4catc,
design=ncsrsvyc10))
Stratified 1 - level Cluster Sampling design (with replacement)
With (84) clusters.
svydesign(strata = ~SESTRAT, id = ~SECLUSTER, weights = ~NCSRWTSH,
data = ncsrc10, nest = T)
Call:
svycoxph(formula = Surv(ageonsetmde, mde) ~ racec + intwage +
sexm + mar3catc + ed4catc, design = ncsrsvyc10)

```

n= 9282, number of events= 1829

	coef	exp(coef)	se(coef)	robust se	z	Pr(> z)
racecHispanic	-0.251413	0.777701	0.136635	0.135175	-1.860	0.06290
racecBlack	-0.481060	0.618128	0.138260	0.149788	-3.212	0.00132
racecWhite	0.078158	1.081294	0.114142	0.118217	0.661	0.50852
intwage	-0.049680	0.951534	0.002170	0.002392	-20.766	< 2e-16
sexm	-0.455350	0.634226	0.049448	0.062540	-7.281	3.31e-13
mar3catcPreviously Married	0.504709	1.656503	0.059794	0.060340	8.364	< 2e-16
mar3catcNever Married	0.081532	1.084948	0.067728	0.089182	0.914	0.36060
ed4catc12	-0.057437	0.944181	0.076388	0.067355	-0.853	0.39380
ed4catc13-15	0.045108	1.046141	0.077950	0.058314	0.774	0.43921
ed4catc16+	-0.091455	0.912603	0.082070	0.063933	-1.430	0.15258

```

racecHispanic      .
racecBlack         **
racecWhite
intwage            ***
sexm               ***
mar3catcPreviously Married ***
mar3catcNever Married
ed4catc12
ed4catc13-15
ed4catc16+

```

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

	exp(coef)	exp(-coef)	lower .95	upper .95
racecHispanic	0.7777	1.2858	0.5967	1.0136
racecBlack	0.6181	1.6178	0.4609	0.8290
racecWhite	1.0813	0.9248	0.8577	1.3632
intwage	0.9515	1.0509	0.9471	0.9560
sexm	0.6342	1.5767	0.5611	0.7169
mar3catcPreviously Married	1.6565	0.6037	1.4717	1.8645
mar3catcNever Married	1.0849	0.9217	0.9110	1.2922
ed4catc12	0.9442	1.0591	0.8274	1.0774
ed4catc13-15	1.0461	0.9559	0.9332	1.1728
ed4catc16+	0.9126	1.0958	0.8051	1.0344

```

Concordance= 0.694 (se = 0.007 )
Likelihood ratio test= NA on 10 df, p=NA
Wald test = 672.5 on 10 df, p=<2e-16
Score (logrank) test = NA on 10 df, p=NA

```

(Note: the likelihood ratio and score tests assume independence of observations within a cluster, the Wald and robust score tests do not).

```
> # No formal test of proportional assumption in R, could try using adjusted model but we have an informal test from kmsvy above
```

```
> # Discrete time logistic using ncsr data in person year format  
> # Read in personyear data, previously set up with multiple records per person (SAS)
```

```
> ncsrpy <- read_sas("P:/ASDA3/Replication SAS/c10_expanded1.sas7bdat")  
> names(ncsrpy)  
 [1] "CASEID"      "DSM_SO"      "MDE_OND"     "SO_OND"     "AGE"  
 [6] "REGION"     "MAR3CAT"     "ED4CAT"     "OBESE6CA"  "NCSRWTSH"  
[11] "NCSRWTLG"   "SEX"        "WKSTAT3C"   "SESTRAT"   "SECLUSTR"  
[16] "ag4cat"     "racecat"    "mde"        "ald"        "sexf"  
[21] "sexm"       "ageonsetmde" "intwage"    "ncsrwtsh100" "pyr"  
[26] "mdetv"
```

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

```
> # Prepare subset of records <= age of onset of mde/censor, needed for model to follow  
> subncsrpy <- subset(ncsrsvypyp1, pyr <= ageonsetmde)
```

```
> # Run DT model  
> 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), design = subncsrpy, family = quasibinomial)
```

```
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 with CIs  
> coef <- exp(coef(ex105_logit))  
> coef  
 (Intercept)          pyr          intwage          sexm factor(ED4CAT)2  
 0.03220851      1.03334155      0.94333508      0.64090809      0.98006512
```

```

factor(ED4CAT)3 factor(ED4CAT)4 factor(racecat)2 factor(racecat)3 factor(racecat)4
  1.09737261    0.98073699    0.78003095    0.63320074    1.07680197
factor(MAR3CAT)2 factor(MAR3CAT)3
  1.63926854    0.96527120
> ci_exp <- exp(confint(ex105_logit))
> ci_exp
      2.5 %    97.5 %
(Intercept)  0.02314675 0.04481787
pyr          1.02898030 1.03772129
intwage      0.93863584 0.94805784
sexm         0.56444816 0.72772525
factor(ED4CAT)2 0.85643412 1.12154295
factor(ED4CAT)3 0.97605128 1.23377396
factor(ED4CAT)4 0.86188838 1.11597404
factor(racecat)2 0.59256897 1.02679741
factor(racecat)3 0.46642055 0.85961731
factor(racecat)4 0.84606830 1.37045966
factor(MAR3CAT)2 1.44747219 1.85647873
factor(MAR3CAT)3 0.80673330 1.15496471
>
>
> # 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), design = subncsrpy, family = quasibinomial(link = cloglog))

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
> coef <- exp(coef(ex105_cloglog))
> coef
      (Intercept)          pyr          intwage          sexm factor(ED4CAT)2
      0.03192411      1.03327451      0.94347978      0.64196511      0.98045363
factor(ED4CAT)3 factor(ED4CAT)4 factor(racecat)2 factor(racecat)3 factor(racecat)4
      1.09675913      0.98097958      0.78080937      0.63439840      1.07652097
factor(MAR3CAT)2 factor(MAR3CAT)3
      1.63691750      0.96514832
> ci_exp <- exp(confint(ex105_cloglog))
> ci_exp
              2.5 %      97.5 %
(Intercept)  0.02297111 0.04436655
pyr          1.02892354 1.03764387
intwage      0.93879546 0.94818746
sexm         0.56561874 0.72861659
factor(ED4CAT)2 0.85722867 1.12139194
factor(ED4CAT)3 0.97599316 1.23246825
factor(ED4CAT)4 0.86252275 1.11570499
factor(racecat)2 0.59364664 1.02698008
factor(racecat)3 0.46772919 0.86045801
factor(racecat)4 0.84647104 1.36909281
factor(MAR3CAT)2 1.44610531 1.85290717
factor(MAR3CAT)3 0.80734624 1.15379404

```