

CHAPTER 10 SURVIVAL ANALYSIS EXAMPLES REPLICATION SPSS/PASW V18  
SURVIVAL ANALYSIS: COX PROPORTIONAL HAZARDS MODEL, KAPLAN-MEIER SURVIVAL CURVES AND DISCRETE TIME LOGISTIC REGRESSION

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 ANALYSIS IN SPSS/PASW V18 COMPLEX SAMPLES MODULE

SPSS/PASW CSCOXREG/CSLOGISTIC/CSORDINAL commands can perform most of the analyses presented in Chapter 10 of ASDA. CSCOXREG performs Cox Proportional Hazards regression and Kaplan Meier survival curves (if run without predictors) with both approaches using a one record per person data set. The CSLOGISTIC command performs discrete time logistic regression while CSORDINAL offers the CLOGLOG link for discrete time logistic regression with the complementary log-log link (using a multiple record per person data set).

Some of the fine points of these procedures are the use of a SUBPOP statement for subpopulation analyses, various output statistics specified on the STATISTICS subcommand, and use of an analysis Plan file for all Complex Samples commands. The plan file should be prepared prior to working with any Complex Samples commands and offers the ability to declare weights and design variables to the program. For matching the reference group to Stata v10.1, we use a reverse coding strategy as this is one way to match the omitted categories of Stata (lowest category is omitted by default). As previously stated, another option would be to use the indicator variable for each level of the categorical variable approach instead of reverse coding.

\* Figure 10.3 Kaplan Meier Survival Curve

\* Complex Samples Cox Regression.

CSCOXREG ageonsetmde

/PLAN FILE='F:\applied\_analysis\_book\SPSS Analysis Examples Replication\Analysis Examples Replication Winter 2010

SPSSv18\ncsr\_p1wt.csaplan'

/VARIABLES STATUS=mde(1) BASELINESTRATA=racecat

/PRINT SAMPLEINFO EVENTINFO

/STATISTICS PARAMETER SE CINTERVAL

/PLOT SURVIVAL CI=NO

/TEST TYPE=F PADJUST=LSD

/CRITERIA MXITER=100 MXSTEP=5 PCONVERGE=[1E-006 RELATIVE] LCONVERGE=[0] TIES=EFRON CILEVEL=95

/SURVIVALMETHOD BASELINE=EFRON CI=LOG

/MISSING CLASSMISSING=EXCLUDE.

**Sample Design Information**

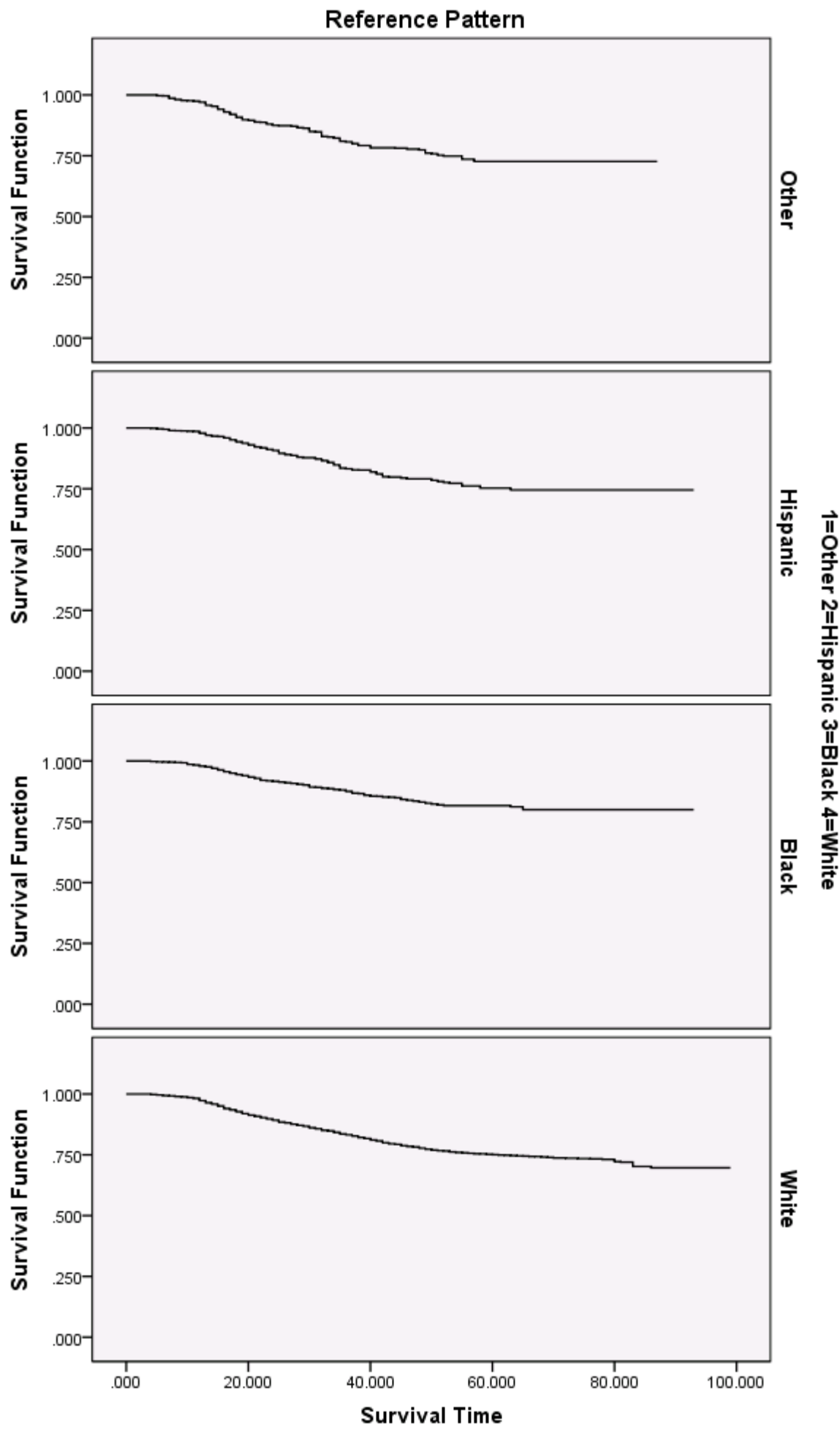
			N
Unweighted Counts	Valid	Subjects	9282
		Cases	9282
		Invalid Cases	0
		Total Cases	9282
Population Subject Size			9.282E3
Stage 1	Strata		42
	Units		84
Sampling Design Degrees of Freedom			42

**Event and Censoring Information**

Baseline Strata	Total		Event		Censored		Censored Percent	
	Unweighted	Weighted	Unweighted	Weighted	Unweighted	Weighted	Unweighted	Weighted
1=Other 2=Hispanic	473	404.334	103	81.677	370	322.657	78.2%	79.8%
3=Black 4=White	883	1006.616	165	164.894	718	841.723	81.3%	83.6%
	1230	1072.792	184	150.990	1046	921.802	85.0%	85.9%
	6696	6798.258	1377	1381.903	5319	5416.355	79.4%	79.7%
Total	9282	9282.000	1829	1779.464	7453	7502.536	80.3%	80.8%

Event Status variable: mde = 1

Survival Function Charts



\*Example for Table 10.2

\* Complex Samples Cox Regression.

```
CSCOXREG ageonsetmde BY revmarcat revracecat reved4cat WITH AGE sexm
/PLAN FILE='F:\applied_analysis_book\SPSS Analysis Examples Replication\Analysis Examples Replication Winter 2010
SPSSv18\ncsr_p1wt.csaplan' /VARIABLES STATUS=mde(1)
/MODEL revmarcat revracecat reved4cat AGE sexm
/PRINT SAMPLEINFO EVENTINFO
/STATISTICS PARAMETER EXP SE CINTERVAL TTEST
/PLOT SURVIVAL CI=NO
/TEST TYPE=F PADJUST=LSD
/CRITERIA MXITER=100 MXSTEP=5 PCONVERGE=[1E-006 RELATIVE] LCONVERGE=[0] TIES=EFRON CILEVEL=95
/SURVIVALMETHOD BASELINE=EFRON CI=LOG
/MISSING CLASSMISSING=EXCLUDE.
```

NOTE: CODES FOR RACECAT 1=OTHER 2=HISPANIC 3=BLACK 4=WHITE, MARCAT 1=MARRIED 2=PREVIOUSLY MARRIED 3=NEVER MARRIED, ED4CAT 1=0-11 2=12 3=13-15 4=16+ YEARS OF EDUCATION, SEXM 1=MALE 0=NON-MALE. REVERSE CODING IS SIMPLY THE REVERSE OF THE ORIGINAL CODES.

**Sample Design Information**

			N
Unweighted Counts	Valid	Subjects	9282
		Cases	9282
		Invalid Cases	0
		Total Cases	9282
Population Subject Size			9.282E3
Stage 1	Strata		42
	Units		84
Sampling Design Degrees of Freedom			42

**Event and Censoring Information**

Total		Event		Censored		Censored Percent	
Unweighted	Weighted	Unweighted	Weighted	Unweighted	Weighted	Unweighted	Weighted
9282	9282.000	1829	1779.464	7453	7502.536	80.3%	80.8%

Event Status variable: mde = 1

**Tests of Model Effects**

Source	df1	df2	Wald F	Sig.
revmarcat	2.000	41.000	34.634	.000
revracecat	3.000	40.000	13.435	.000
reved4cat	3.000	40.000	1.975	.133
AGE	1.000	42.000	431.229	.000
sexm	1.000	42.000	53.012	.000

Survival Time Variable: ageonsetmde

Event Status Variable: mde = 1

Model: revmarcat, revracecat, reved4cat, AGE, sexm

Parameter Estimates<sup>b</sup>

Parameter	B	Std. Error	95% Confidence Interval		Hypothesis Test			Exp(B)	95% Confidence Interval for Exp(B)	
			Lower	Upper	t	df	Sig.		Lower	Upper
[revmarcat=1.0000]	.082	.089	-.098	.262	.914	42.000	.366	1.085	.906	1.299
[revmarcat=2.0000]	.505	.060	.383	.626	8.364	42.000	.000	1.657	1.467	1.871
[revmarcat=3.0000]	.000 <sup>a</sup>	.	.	.	.	.	.	1.000	.	.
[revracecat=1.0000]	.078	.118	-.160	.317	.661	42.000	.512	1.081	.852	1.373
[revracecat=2.0000]	-.481	.150	-.783	-.179	-3.212	42.000	.003	.618	.457	.836
[revracecat=3.0000]	-.251	.135	-.524	.021	-1.860	42.000	.070	.778	.592	1.022
[revracecat=4.0000]	.000 <sup>a</sup>	.	.	.	.	.	.	1.000	.	.
[reved4cat=1.0000]	-.091	.064	-.220	.038	-1.430	42.000	.160	.913	.802	1.038
[reved4cat=2.0000]	.045	.058	-.073	.163	.774	42.000	.444	1.046	.930	1.177
[reved4cat=3.0000]	-.057	.067	-.193	.078	-.853	42.000	.399	.944	.824	1.082
[reved4cat=4.0000]	.000 <sup>a</sup>	.	.	.	.	.	.	1.000	.	.
AGE	-.050	.002	-.055	-.045	-20.766	42.000	.000	.952	.947	.956
sexm	-.455	.063	-.582	-.329	-7.281	42.000	.000	.634	.559	.720

Survival Time Variable: ageonsetmde

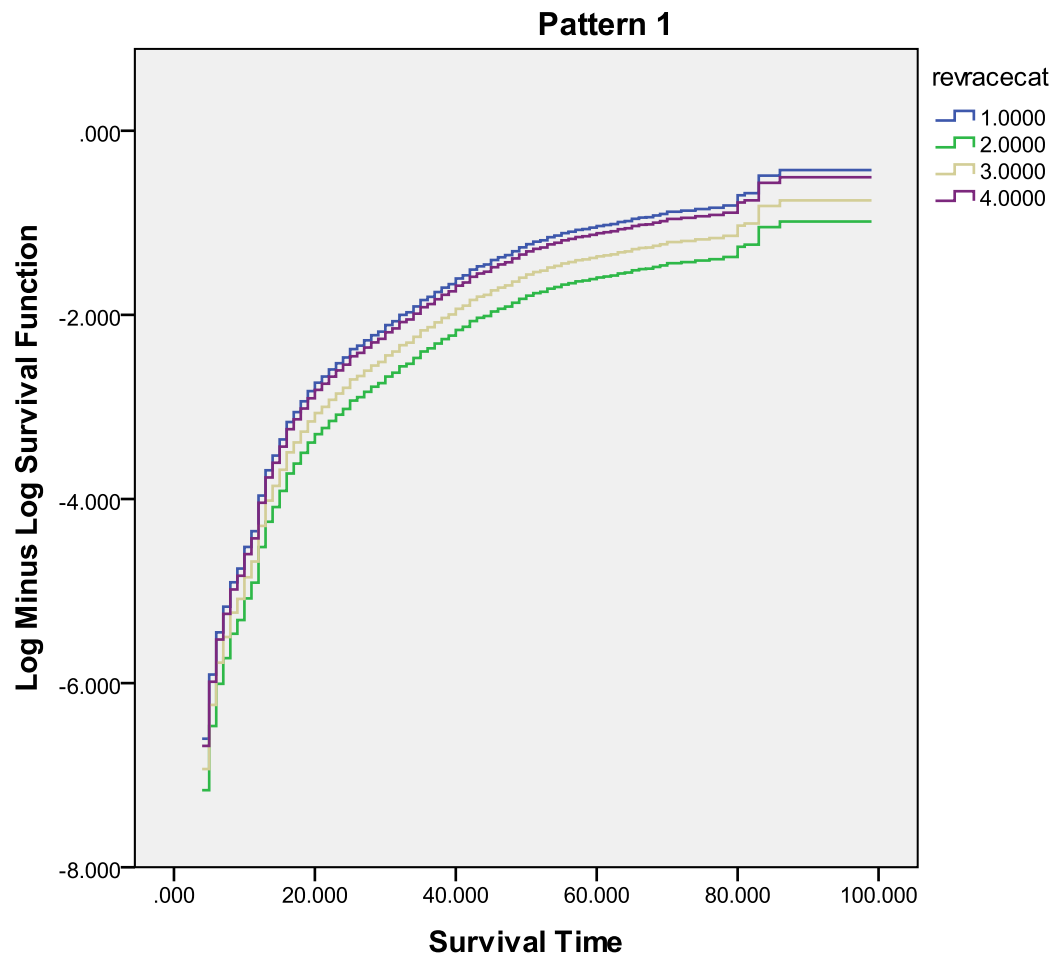
Event Status Variable: mde = 1

Model: revmarcat, revracecat, reved4cat, AGE, sexm

a. Set to zero because this parameter is redundant.

b. Tie breaking method: Efron

```
*test proportional hazards assumption for the fitted Cox model
* Complex Samples Cox Regression.
CSCOXREG ageonsetmde BY revmarcat revracecat reved4cat WITH AGE sexm
  /PLAN FILE='F:\applied_analysis_book\SPSS Analysis Examples Replication\Analysis Examples Replication Winter 2010
SPSSv18\ncsr_p1wt.csaplan' /VARIABLES STATUS=mde(1)
/MODEL revmarcat revracecat reved4cat AGE sexm
/PRINT SAMPLEINFO EVENTINFO
/STATISTICS PARAMETER EXP SE CINTERVAL TTEST
/PLOT LML CI=NO
/PATTERN BY revracecat
/TEST TYPE=F PADJUST=LSD
/CRITERIA MXITER=100 MXSTEP=5 PCONVERGE=[1E-006 RELATIVE] LCONVERGE=[0] TIES=EFRON CILEVEL=95
/SURVIVALMETHOD BASELINE=EFRON CI=LOG
/MISSING CLASSMISSING=EXCLUDE.
```



\*Discrete Time Logistic Regression  
 \*person year file is already created

```
USE ALL.
COMPUTE filter_$=(pyr = ageonsetmde ).
VARIABLE LABEL filter_$ 'pyr = ageonsetmde (FILTER)'.
VALUE LABELS filter_$ 0 'Not Selected' 1 'Selected'.
FORMAT filter_$ (f1.0).
FILTER BY filter_$.
EXECUTE.
```

```
* Complex Samples Logistic Regression.
* Discrete Time Logistic Regression NCS-R data.
CSLOGISTIC mdetv(LOW) BY revedcat revmarcat revracecat WITH pyr sexm intwage
/PLAN FILE='F:\applied_analysis_book\SPSS Analysis Examples Replication\Analysis Examples Replication Winter 2010
SPSSv18\ncsr_plwt.csaplan'
/MODEL revedcat revmarcat revracecat pyr sexm intwage
/INTERCEPT INCLUDE=YES SHOW=YES
/STATISTICS PARAMETER EXP SE CINTERVAL TTEST
/TEST TYPE=F PADJUST=LSD
/MISSING CLASSMISSING=EXCLUDE
/CRITERIA MXITER=100 MXSTEP=5 PCONVERGE=[1E-006 RELATIVE] LCONVERGE=[0] CHKSEP=20 CILEVEL=95
/PRINT SUMMARY SAMPLEINFO.
```

**Sample Design Information**

		N
Unweighted Cases	Valid	385696
	Invalid	0
	Total	385696
Population Size		3.869E5
Stage 1	Strata	42
	Units	84
Sampling Design Degrees of Freedom		42

**Pseudo R Squares**

Cox and Snell	.003
Nagelkerke	.047
McFadden	.046

Dependent Variable: mdetv

(reference category = 0)

Model: (Intercept), revedcat,

revmarcat, revracecat, pyr,

sexm, intwage

**Tests of Model Effects**

Source	df1	df2	Wald F	Sig.
(Corrected Model)	11.000	32.000	53.633	.000
(Intercept)	1.000	42.000	1.494E3	.000
revedcat	3.000	40.000	1.864	.151
revmarcat	2.000	41.000	34.351	.000
revracecat	3.000	40.000	11.963	.000
pyr	1.000	42.000	250.133	.000
sexm	1.000	42.000	51.010	.000
intwage	1.000	42.000	567.546	.000

**Tests of Model Effects**

Source	df1	df2	Wald F	Sig.
(Corrected Model)	11.000	32.000	53.633	.000
(Intercept)	1.000	42.000	1.494E3	.000
revedcat	3.000	40.000	1.864	.151
revmarcat	2.000	41.000	34.351	.000
revracecat	3.000	40.000	11.963	.000
pyr	1.000	42.000	250.133	.000
sexm	1.000	42.000	51.010	.000
intwage	1.000	42.000	567.546	.000

Dependent Variable: mdetv (reference category = 0)

Model: (Intercept), revedcat, revmarcat, revracecat, pyr, sexm, intwage

**Parameter Estimates**

mdetv	Parameter	B	Std. Error	95% Confidence Interval		Hypothesis Test			Exp(B)	95% Confidence Interval for Exp(B)	
				Lower	Upper	t	df	Sig.		Lower	Upper
1	(Intercept)	-3.436	.162	-3.762	-3.109	-21.209	42.000	.000	.032	.023	.045
	[revedcat=1.0000]	-.019	.063	-.147	.108	-.307	42.000	.760	.981	.863	1.114
	[revedcat=2.0000]	.093	.057	-.023	.209	1.618	42.000	.113	1.097	.977	1.232
	[revedcat=3.0000]	-.020	.066	-.154	.113	-.305	42.000	.762	.980	.858	1.120
	[revedcat=4.0000]	.000 <sup>a</sup>	.	.	.	.	.	.	1.000	.	.
	[revmarcat=1.0000]	-.035	.088	-.213	.142	-.402	42.000	.690	.965	.808	1.153
	[revmarcat=2.0000]	.494	.061	.371	.617	8.101	42.000	.000	1.639	1.449	1.854
	[revmarcat=3.0000]	.000 <sup>a</sup>	.	.	.	.	.	.	1.000	.	.
	[revracecat=1.0000]	.074	.118	-.165	.313	.626	42.000	.535	1.077	.848	1.367
	[revracecat=2.0000]	-.457	.150	-.759	-.154	-3.049	42.000	.004	.633	.468	.857
	[revracecat=3.0000]	-.248	.135	-.520	.024	-1.843	42.000	.072	.780	.594	1.024
	[revracecat=4.0000]	.000 <sup>a</sup>	.	.	.	.	.	.	1.000	.	.
	pyr	.033	.002	.029	.037	15.816	42.000	.000	1.033	1.029	1.038
	sexm	-.445	.062	-.571	-.319	-7.142	42.000	.000	.641	.565	.727
	intwage	-.058	.002	-.063	-.053	-23.823	42.000	.000	.943	.939	.948

Dependent Variable: mdetv (reference category = 0)

Model: (Intercept), revedcat, revmarcat, revracecat, pyr, sexm, intwage

a. Set to zero because this parameter is redundant.



\* Complex Samples Ordinal Regression.

\* CLOGLOG link, note that signs differ but coefficients match the Stata output, see documentation for details

\* Repeat model but run with cloglog link, note that Odds Ratios are not provided with the Cloglog link, could be calculated as the exponent(parameter).

```
CSORDINAL mdetv (descending) BY revedcat revmarcat revracecat WITH sexm pyr intwage
/PLAN FILE='F:\applied_analysis_book\SPSS Analysis Examples Replication\Analysis Examples Replication Winter 2010
SPSSv18\ncsr_plwt.csaplan'
/LINK FUNCTION=CLOGLOG
/MODEL revedcat revmarcat revracecat sexm pyr intwage
/STATISTICS PARAMETER SE CINTERVAL TTEST
/TEST TYPE=F PADJUST=LSD
/MISSING CLASSMISSING=EXCLUDE
/CRITERIA MXITER=100 MXSTEP=5 PCONVERGE=[1E-006 RELATIVE] LCONVERGE=[0] METHOD=NEWTON CHKSEP=20 CILEVEL=95
/PRINT SUMMARY CLASSTABLE VARIABLEINFO SAMPLEINFO.
```

#### Sample Design Information

		N
Unweighted Cases	Valid	385696
	Invalid	0
	Total	385696
Population Size		3.869E5
Stage 1	Strata	42
	Units	84
Sampling Design Degrees of Freedom		42

#### Categorical Variable Information

		Weighted Count	Weighted Percent
mdetv <sup>a</sup>	1	1779.464	.5%
	0	385086.583	99.5%
Yrs Education	1=16+ 2=13-15 3=12 4=0-11	92415.979	23.9%
	2.0000	97874.347	25.3%
	3.0000	126116.110	32.6%
	4.0000	70459.611	18.2%
1=Never Married	1.0000	57924.423	15.0%
	2=Previously Married	98110.391	25.4%
	3=Married	230831.233	59.7%
1=White 2=Black 3=Hispanic	1.0000	291968.896	75.5%
	4=Other	42856.612	11.1%
	3.0000	36695.864	9.5%
	4.0000	15344.675	4.0%
Population Size		386866.047	100.0%

a. Dependent variable values are sorted in descending order.

#### Covariate Information

	Mean
sexm	.48
pyr	25.678147
intwage	51.952937

#### Pseudo R Squares

Cox and Snell	.003
Nagelkerke	.047
McFadden	.046

**Pseudo R Squares**

Cox and Snell	.003
Nagelkerke	.047
McFadden	.046

Dependent Variable: mdetv

(Descending)

Model: (Threshold), revedcat,

revmarcat, revracecat, sexm,

pyr, intwage

Link function: Complementary

log-log

**Tests of Model Effects**

Source	df1	df2	Wald F	Sig.
revedcat	3.000	40.000	1.850	.154
revmarcat	2.000	41.000	34.428	.000
revracecat	3.000	40.000	11.939	.000
sexm	1.000	42.000	50.968	.000
pyr	1.000	42.000	250.294	.000
intwage	1.000	42.000	568.002	.000

Dependent Variable: mdetv (Descending)

Model: (Threshold), revedcat, revmarcat, revracecat, sexm, pyr,

intwage

Link function: Complementary log-log

**Parameter Estimates**

Parameter	B	Std. Error	95% Confidence Interval		Hypothesis Test		
			Lower	Upper	t	df	Sig.
Threshold [mdetv=1]	-3.444	.161	-3.770	-3.119	-21.344	42.000	.000
Regression [revedcat=1.0000]	.019	.063	-.108	.147	.304	42.000	.762
[revedcat=2.0000]	-.092	.057	-.208	.023	-1.614	42.000	.114
[revedcat=3.0000]	.020	.066	-.113	.153	.300	42.000	.766
[revedcat=4.0000]	.000 <sup>a</sup>	.	.	.	.	.	.
[revmarcat=1.0000]	.035	.088	-.141	.212	.405	42.000	.687
[revmarcat=2.0000]	-.493	.061	-.615	-.370	-8.107	42.000	.000
[revmarcat=3.0000]	.000 <sup>a</sup>	.	.	.	.	.	.
[revracecat=1.0000]	-.074	.118	-.312	.164	-.626	42.000	.535
[revracecat=2.0000]	.455	.149	.153	.757	3.045	42.000	.004
[revracecat=3.0000]	.247	.134	-.024	.519	1.841	42.000	.073
[revracecat=4.0000]	.000 <sup>a</sup>	.	.	.	.	.	.
sexm	.443	.062	.318	.569	7.139	42.000	.000
pyr	-.033	.002	-.037	-.029	-15.821	42.000	.000
intwage	.058	.002	.053	.063	23.833	42.000	.000

Dependent Variable: mdetv (Descending)

Model: (Threshold), revedcat, revmarcat, revracecat, sexm, pyr, intwage

Link function: Complementary log-log

a. Set to zero because this parameter is redundant.

**Classification**

Observed	Predicted		
	0	1	Percent Correct
0	3.851E5	.000	100.0%
1	1.779E3	.000	.0%
Overall Percent	100.0%	.0%	99.5%

Dependent Variable: mdetv (Descending)

Model: (Threshold), revedcat, revmarcat, revracecat, sexm, pyr, intwage

Link function: Complementary log-log