

SUDAAN Analysis Example Replication C10

```
* Sudaan Analysis Examples Replication for ASDA 2nd Edition
* Berglund April 2017
* Chapter 10 ;
libname ncsr "P:\ASDA 2\Data sets\ncsr\" ;
```

```
options nodate nonumber ;
```

```
data c10_ncsr ;
  set ncsr.ncsr_sub_5apr2017 ;
* prepare variables for analysis ;
  if mde=1 then ageonsetmde=mde_ond ; else ageonsetmde=age ;
  intwage=age ;
  ncsrwtsh100=ncsrwtsh*100 ;
run ;
proc sort ;
  by sestrat seclustr ;
run ;
```

```
title "Distribution of Age of Onset of MDE or Censor" ;
proc freq ;
  tables ageonsetmde ;
run ;
```

```
title "Section 10.3.3 KM Example, Total Sample " ;
proc kapmeier filetype=sas data=c10_ncsr ;
  nest sestrat seclustr ; weight ncsrwtsh ;
  event mde ;
  time ageonsetmde ;
  output / kapmeier=all filename=work.kmcurve filetype=sas replace ;
  setenv decwidth=4 ;
run ;
proc format ;
  value rf 1='Other/Asian' 2='Hispanic/Mexican' 3='Black' 4='White' ;
run ;
proc print data=work.kmcurve ;
  where ageonsetmde in (10,20,30,40,50,60,70) ;
  format _numeric_ 7.4 ;
run ;
```

```
title "Section 10.3.3. KM Example, Stratified by Race" ;
proc kapmeier filetype=sas data=c10_ncsr ;
  nest sestrat seclustr ; weight ncsrwtsh ;
  class racecat ; event mde ;
  time ageonsetmde ;
  strhaz racecat ;
  output / kapmeier=all filename=work.kmcurve_race filetype=sas replace ;
  setenv decwidth=4 ;
run ;
proc format ;
  value rf 1='Other/Asian' 2='Hispanic/Mexican' 3='Black' 4='White' ;
run ;
proc print data=work.kmcurve_race ;
  where ageonsetmde in (10,20,30,40,50,60,70) ;
  format _numeric_ 7.4 ;
  format racecat rf. ;
run ;
```

* Note, Sudaan does not include graphical ability but this can be done with the output data set from PROC KAPMEIER with SAS or similar ;

```
title " 10.4.5 Example: Fitting a Cox Proportional Hazards Model to Complex Sample Survey Data, Table 10.2 " ;
proc survival filetype=sas data=c10_ncsr ;
  nest sestrat seclustr ; weight ncsrwtsh ;
  class sex mar3cat ed4cat racecat / nofreq ;
  *rformat ed4cat edf. ; *rformat mar3cat marf. ;
  reflevel sex=2 mar3cat=1 ed4cat=1 racecat=1 ;
  event mde ;
  model ageonsetmde = intwage sex mar3cat ed4cat racecat ;
  setenv decwidth=4 ;
run ;
```

```
title "10.5.5 Fitting a Discrete Time Model to Complex Sample Survey Data" ;
data c10_expanded ;
  set c10_ncsr ;
```

```

* prepare data for model, shape into "long" file ;
do pyr= 1 to intwage ;
    output ;
end ;
run ;

data c10_expanded1 ;
set c10_expanded ;
if pyr=mde_ond then mdetv=1 ; else mdetv=0 ;
run ;
proc sort ;
by sestrat seclustr ;
run ;

title "Discrete Time Logistic Regression: Table 10.5" ;
proc rlogist data=c10_expanded1 filetype=sas ;
nest sestrat seclustr ;
weight ncsrwtsh ;
class mar3cat ed4cat racecat sex / nofreq ;
reflevel mar3cat=1 ed4cat=1 racecat=1 sex=2 ;
subpopn pyr <= ageonsetmde ;
model mdetv = pyr intwage sex ed4cat racecat mar3cat ;
setenv decwidth=3 ;
test adjwaldf ;
print / style=nchs ;
run ;

* Note, no CLOGLOG option in Sudaan ;

```

Distribution of Age of Onset of MDE or Censor

The FREQ Procedure

ageonsetmde	Frequency	Percent	Cumulative Frequency	Cumulative Percent
4	20	0.22	20	0.22
5	18	0.19	38	0.41
6	19	0.20	57	0.61
7	19	0.20	76	0.82
8	23	0.25	99	1.07
9	16	0.17	115	1.24
10	34	0.37	149	1.61
11	28	0.30	177	1.91
12	76	0.82	253	2.73
13	70	0.75	323	3.48
14	54	0.58	377	4.06
15	66	0.71	443	4.77
16	88	0.95	531	5.72
17	64	0.69	595	6.41
18	214	2.31	809	8.72
19	212	2.28	1021	11.00
20	222	2.39	1243	13.39
21	200	2.15	1443	15.55
22	195	2.10	1638	17.65
23	184	1.98	1822	19.63
24	176	1.90	1998	21.53
25	203	2.19	2201	23.71
26	159	1.71	2360	25.43
27	194	2.09	2554	27.52
28	162	1.75	2716	29.26
29	152	1.64	2868	30.90
30	233	2.51	3101	33.41
31	148	1.59	3249	35.00
32	186	2.00	3435	37.01
33	159	1.71	3594	38.72
34	180	1.94	3774	40.66
35	194	2.09	3968	42.75
36	171	1.84	4139	44.59
37	182	1.96	4321	46.55
38	215	2.32	4536	48.87
39	153	1.65	4689	50.52
40	209	2.25	4898	52.77
41	158	1.70	5056	54.47
42	194	2.09	5250	56.56
43	193	2.08	5443	58.64
44	162	1.75	5605	60.39
45	151	1.63	5756	62.01
46	126	1.36	5882	63.37
47	160	1.72	6042	65.09
48	142	1.53	6184	66.62
49	160	1.72	6344	68.35
50	150	1.62	6494	69.96
51	124	1.34	6618	71.30
52	141	1.52	6759	72.82
53	122	1.31	6881	74.13
54	115	1.24	6996	75.37
55	94	1.01	7090	76.38
56	121	1.30	7211	77.69
57	100	1.08	7311	78.77
58	112	1.21	7423	79.97
59	96	1.03	7519	81.01
60	103	1.11	7622	82.12
61	76	0.82	7698	82.93
62	82	0.88	7780	83.82
63	74	0.80	7854	84.62

Distribution of Age of Onset of MDE or Censor

The FREQ Procedure

ageonsetmde	Frequency	Percent	Cumulative Frequency	Cumulative Percent
64	90	0.97	7944	85.59
65	79	0.85	8023	86.44
66	70	0.75	8093	87.19
67	71	0.76	8164	87.96
68	90	0.97	8254	88.92
69	70	0.75	8324	89.68
70	64	0.69	8388	90.37
71	75	0.81	8463	91.18
72	63	0.68	8526	91.86
73	65	0.70	8591	92.56
74	75	0.81	8666	93.36
75	59	0.64	8725	94.00
76	75	0.81	8800	94.81
77	57	0.61	8857	95.42
78	64	0.69	8921	96.11
79	45	0.48	8966	96.60
80	54	0.58	9020	97.18
81	54	0.58	9074	97.76
82	32	0.34	9106	98.10
83	30	0.32	9136	98.43
84	32	0.34	9168	98.77
85	17	0.18	9185	98.95
86	19	0.20	9204	99.16
87	19	0.20	9223	99.36
88	13	0.14	9236	99.50
89	10	0.11	9246	99.61
90	15	0.16	9261	99.77
91	6	0.06	9267	99.84
92	4	0.04	9271	99.88
93	5	0.05	9276	99.94
94	2	0.02	9278	99.96
95	1	0.01	9279	99.97
98	2	0.02	9281	99.99
99	1	0.01	9282	100.00

Section 10.3.3 KM Example, Total Sample

S U D A A N
 Software for the Statistical Analysis of Correlated Data
 Copyright Research Triangle Institute June 2013
 Release 11.0.1

DESIGN SUMMARY: Variances will be computed using the Taylor Linearization Method, Assuming a With Replacement (WR) Design

Sample Weight: NCSRWTSH
 Stratification Variables(s): SESTRAT
 Primary Sampling Unit: SECLUSTR

Number of observations read : 9282 Weighted count: 9282
 Observations used in the analysis : 9282 Weighted count: 9282
 Denominator degrees of freedom : 42

Date: 05-18-2017
 Time: 14:27:31

SUDAAN

Page: 1
 Table: 1

Summary of Event Values
 by: Major Depressive Episode 1=Yes 0=No.

```

-----
Major Depressive
Episode 1=Yes 0=No      Frequency      Weighted Sum
-----
Censored                7453.000      7502.536
Non-Censored            1829.000      1779.464
-----
    
```

Section 10.3.3 KM Example, Total Sample

Obs	PROCNUM	AGEONSETMDE	KM	SEKM	LOWKM	UPKM
7	14.0000	10.0000	0.9850	0.0013	0.9821	0.9875
17	14.0000	20.0000	0.9165	0.0038	0.9084	0.9239
27	14.0000	30.0000	0.8659	0.0050	0.8554	0.8757
37	14.0000	40.0000	0.8155	0.0054	0.8043	0.8262
47	14.0000	50.0000	0.7760	0.0058	0.7641	0.7874
57	14.0000	60.0000	0.7568	0.0060	0.7445	0.7686
67	14.0000	70.0000	0.7438	0.0066	0.7303	0.7568

Section 10.3.3. KM Example, Stratified by Race

S U D A A N
 Software for the Statistical Analysis of Correlated Data
 Copyright Research Triangle Institute June 2013
 Release 11.0.1

DESIGN SUMMARY: Variances will be computed using the Taylor Linearization Method, Assuming a With Replacement (WR) Design

Sample Weight: NCSRWTSH
 Stratification Variables(s): SESTRAT
 Primary Sampling Unit: SECLUSTR

Number of observations read : 9282 Weighted count: 9282
 Observations used in the analysis : 9282 Weighted count: 9282
 Denominator degrees of freedom : 42

Date: 05-18-2017
 Time: 14:27:32

SUDAAN

Page: 1
 Table: 1

Summary of Event Values

by: Major Depressive Episode 1=Yes 0=No, Race 1=Other/Asian 2=Hisp/Mexican 3=Black 4=White.

```
-----
```

Major Depressive Episode 1=Yes 0=No			
Race			
1=Other/Asian			
2=Hisp/Mexican			
3=Black 4=White			
	Frequency	Weighted Sum	

Censored			
1	370.000	322.657	
2	718.000	841.723	
3	1046.000	921.802	
4	5319.000	5416.355	
Non-Censored			
1	103.000	81.677	
2	165.000	164.894	
3	184.000	150.990	
4	1377.000	1381.903	

Frequencies and Values for CLASS Variables
 by: Race 1=Other/Asian 2=Hisp/Mexican 3=Black 4=White.

```

-----
Race
 1=Other/A-
sian
 2=Hisp/Me-
xican
 3=Black
 4=White
      Frequency      Value
-----
Ordered
  Position:
 1              473          1
Ordered
  Position:
 2              883          2
Ordered
  Position:
 3             1230          3
Ordered
  Position:
 4             6696          4
-----
  
```

Section 10.3.3. KM Example, Stratified by Race

Obs	PROCNUM	AGEONSETMDE	KM	SEKM	LOWKM	UPKM	RACECAT
6	14.0000	10.0000	0.9763	0.0087	0.9505	0.9887	Other/Asian
16	14.0000	20.0000	0.8954	0.0193	0.8490	0.9281	Other/Asian
25	14.0000	30.0000	0.8495	0.0199	0.8042	0.8851	Other/Asian
34	14.0000	40.0000	0.7821	0.0240	0.7289	0.8262	Other/Asian
39	14.0000	50.0000	0.7577	0.0244	0.7041	0.8030	Other/Asian
50	14.0000	10.0000	0.9867	0.0041	0.9752	0.9929	Hisp/Mexican
60	14.0000	20.0000	0.9303	0.0072	0.9141	0.9435	Hisp/Mexican
78	14.0000	40.0000	0.8191	0.0188	0.7775	0.8536	Hisp/Mexican
84	14.0000	50.0000	0.7847	0.0218	0.7367	0.8251	Hisp/Mexican
97	14.0000	10.0000	0.9860	0.0027	0.9794	0.9905	Black
107	14.0000	20.0000	0.9335	0.0075	0.9166	0.9471	Black
117	14.0000	30.0000	0.8927	0.0122	0.8653	0.9149	Black
127	14.0000	40.0000	0.8554	0.0156	0.8206	0.8840	Black
137	14.0000	50.0000	0.8227	0.0159	0.7879	0.8523	Black
148	14.0000	10.0000	0.9851	0.0017	0.9812	0.9883	White
158	14.0000	20.0000	0.9130	0.0042	0.9040	0.9212	White
168	14.0000	30.0000	0.8610	0.0056	0.8492	0.8720	White
178	14.0000	40.0000	0.8108	0.0055	0.7995	0.8215	White
188	14.0000	50.0000	0.7694	0.0053	0.7584	0.7799	White
198	14.0000	60.0000	0.7504	0.0061	0.7378	0.7625	White
208	14.0000	70.0000	0.7371	0.0065	0.7236	0.7501	White

* NOTE: no graphics directly from SUDAAN, use SAS or other tool.

S U D A A N
 Software for the Statistical Analysis of Correlated Data
 Copyright Research Triangle Institute June 2013
 Release 11.0.1

DESIGN SUMMARY: Variances will be computed using the Taylor Linearization Method, Assuming a With Replacement (WR) Design

Sample Weight: NCSRWTSH
 Stratification Variables(s): SESTRAT
 Primary Sampling Unit: SECLUSTR

NOTE: Using a default start time of -10000000000 for all records

Number of observations read : 9282 Weighted count: 9282
 Observations used in the analysis : 9282 Weighted count: 9282
 Denominator degrees of freedom : 42

Maximum number of estimable parameters for the model is 10

Date: 05-18-2017
 Time: 14:27:34

SUDAAN

Page: 1
 Table: 1

Summary of Event Values
 by: Major Depressive Episode 1=Yes 0=No.

```
-----
```

Major Depressive Episode 1=Yes 0=No	Frequency	Weighted Sum
Censored	7453.000	7502.536
Non-Censored	1829.000	1779.464

```
-----
```

SURVIVAL has converged to a solution in 5 iterations.

-2 * Normalized Log-Likelihood with Beta(s) = 0 : 31216.13
 -2 * Normalized Log-Likelihood Full Model : 30350.14
 Approximate Chi-Square (-2 * Log-L Ratio) : 865.98
 Degrees of Freedom : 10
 Approximate P-Value : 0.00

Note: The approximate Chi-Square is not adjusted for clustering.
 Refer to hypothesis test table for adjusted test.

Date: 05-18-2017
 Time: 14:27:34

SUDAAN

Page: 2
 Table: 1

Variance Estimation Method: Taylor Series (WR)
 Dependent Variable: AGEONSETMDE
 Censoring Variable: MDE: Major Depressive Episode 1=Yes 0=No
 Ties Handling: EFRON
 by: Independent Variables and Effects.

Independent Variables and Effects	Beta Coeff.	SE Beta	Lower 95% Limit Beta	Upper 95% Limit Beta	T-Test B=0	P-value T-Test B=0
INTWAGE	-0.0497	0.0024	-0.0545	-0.0449	-20.7660	0.0000
Sex 1=Male 2=Female						
1	-0.4554	0.0625	-0.5816	-0.3291	-7.2810	0.0000
2	0.0000	0.0000	0.0000	0.0000	.	.
Marital Status						
1=Married						
2=Previously Married						
3=Never Married						
1	0.0000	0.0000	0.0000	0.0000	.	.
2	0.5047	0.0603	0.3829	0.6265	8.3644	0.0000
3	0.0815	0.0892	-0.0984	0.2615	0.9142	0.3658
Education 1=0-11						
2=12 3=13-15 4=16+ Yrs						
1	0.0000	0.0000	0.0000	0.0000	.	.
2	-0.0574	0.0674	-0.1934	0.0785	-0.8528	0.3986
3	0.0451	0.0583	-0.0726	0.1628	0.7735	0.4435
4	-0.0915	0.0639	-0.2205	0.0376	-1.4305	0.1600
Race 1=Other/Asian						
2=Hispanic/Mexican						
3=Black 4=White						
1	0.0000	0.0000	0.0000	0.0000	.	.
2	-0.2514	0.1352	-0.5242	0.0214	-1.8599	0.0699
3	-0.4811	0.1498	-0.7833	-0.1788	-3.2116	0.0025
4	0.0782	0.1182	-0.1604	0.3167	0.6611	0.5121

Date: 05-18-2017
 Time: 14:27:34

SUDAAN

Page: 3
 Table: 1

Variance Estimation Method: Taylor Series (WR)
 Dependent Variable: AGEONSETMDE
 Censoring Variable: MDE: Major Depressive Episode 1=Yes 0=No
 Ties Handling: EFRON
 by: Contrast.

Contrast	Degrees of Freedom	Wald F	P-value Wald F
OVERALL MODEL	10.0000	67.2484	0.0000
INTWAGE	1.0000	431.2288	0.0000
SEX	1.0000	53.0124	0.0000
MAR3CAT	2.0000	35.4786	0.0000
ED4CAT	3.0000	2.0735	0.1182
RACECAT	3.0000	14.1063	0.0000

Variance Estimation Method: Taylor Series (WR)
 Dependent Variable: AGEONSETMDE
 Censoring Variable: MDE: Major Depressive Episode 1=Yes 0=No
 Ties Handling: EFRON
 by: Independent Variables and Effects.

Independent Variables and Effects	Hazards Ratio	Lower 95% Limit	Upper 95% Limit
INTWAGE	0.9515	0.9470	0.9561
Sex 1=Male 2=Female			
1	0.6342	0.5590	0.7195
2	1.0000	1.0000	1.0000
Marital Status			
1=Married			
2=Previously Married			
3=Never Married			
1	1.0000	1.0000	1.0000
2	1.6565	1.4666	1.8710
3	1.0849	0.9062	1.2989
Education 1=0-11			
2=12 3=13-15 4=16+			
Yrs			
1	1.0000	1.0000	1.0000
2	0.9442	0.8242	1.0817
3	1.0461	0.9300	1.1768
4	0.9126	0.8021	1.0383
Race 1=Other/Asian			
2=Hispanic/Mexican			
3=Black 4=White			
1	1.0000	1.0000	1.0000
2	0.7777	0.5920	1.0216
3	0.6181	0.4569	0.8363
4	1.0813	0.8518	1.3726

Discrete Time Logistic Regression: Table 10.5

S U D A A N
Software for the Statistical Analysis of Correlated Data
Copyright Research Triangle Institute June 2013
Release 11.0.1

DESIGN SUMMARY: Variances will be computed using the Taylor Linearization Method, Assuming a With Replacement (WR) Design

Sample Weight: NCSRWTSH
Stratification Variables(s): SESTRAT
Primary Sampling Unit: SECLUSTER

Number of zero responses : 383867
Number of non-zero responses : 1829

Independence parameters have converged in 7 iterations.

Number of observations read	: 415219	Weighted count:	415607
Observations in subpopulation	: 385696	Weighted count:	386866
Observations used in the analysis	: 385696	Weighted count:	386866
Denominator degrees of freedom	: 42		

Maximum number of estimable parameters for the model is 12

File C10_EXPANDED1 contains 84 Clusters
84 clusters were used to fit the model
Maximum cluster size is 9268 records
Minimum cluster size is 1168 records

Sample and Population Counts for Response Variable MDETV
Based on observations used in the analysis
0: Sample Count 383867 Population Count 385087
1: Sample Count 1829 Population Count 1779

R-Square for dependent variable MDETV (Cox & Snell, 1989): 0.002700

-2 * Normalized Log-Likelihood with Intercepts Only	: 22635.38
-2 * Normalized Log-Likelihood Full Model	: 21592.41
Approximate Chi-Square (-2 * Log-L Ratio)	: 1042.97
Degrees of Freedom	: 11

Note: The approximate Chi-Square is not adjusted for clustering.
Refer to hypothesis test table for adjusted test.

Variance Estimation Method: Taylor Series (WR)
 SE Method: Robust (Binder, 1983)
 Working Correlations: Independent
 Link Function: Logit
 Response variable MDETV: MDETV
 For Subpopulation: PYR <= AGEONSETMDE
 by: Independent Variables and Effects.

Independent Variables and Effects	Beta Coeff.	SE Beta	Lower 95% Limit Beta	Upper 95% Limit Beta	T-Test B=0	P-value T-Test B=0
Intercept	-3.436	0.162	-3.762	-3.109	-21.209	0.000
PYR	0.033	0.002	0.029	0.037	15.816	0.000
INTWAGE	-0.058	0.002	-0.063	-0.053	-23.823	0.000
Sex 1=Male 2=Female						
1	-0.445	0.062	-0.571	-0.319	-7.142	0.000
2	0.000	0.000	0.000	0.000	.	.
Education 1=0-11 2=12 3=13-15 4=16+ Yrs						
1	0.000	0.000	0.000	0.000	.	.
2	-0.020	0.066	-0.154	0.113	-0.305	0.762
3	0.093	0.057	-0.023	0.209	1.618	0.113
4	-0.019	0.063	-0.147	0.108	-0.307	0.760
Race 1=Other/Asian 2=Hisp/Mexican 3=Black 4=White						
1	0.000	0.000	0.000	0.000	.	.
2	-0.248	0.135	-0.520	0.024	-1.843	0.072
3	-0.457	0.150	-0.759	-0.154	-3.049	0.004
4	0.074	0.118	-0.165	0.313	0.626	0.535
Marital Status 1=Married 2=Previously Married 3=Never Married						
1	0.000	0.000	0.000	0.000	.	.
2	0.494	0.061	0.371	0.617	8.101	0.000
3	-0.035	0.088	-0.213	0.142	-0.402	0.690

Date: 05-18-2017
 Time: 14:27:37

SUDAAN

Page: 2
 Table: 1

Variance Estimation Method: Taylor Series (WR)
 SE Method: Robust (Binder, 1983)
 Working Correlations: Independent
 Link Function: Logit
 Response variable MDETV: MDETV
 For Subpopulation: PYR <= AGEONSETMDE
 by: Contrast.

Contrast	Degrees of Freedom	P-value	
		Adj Wald F	Adj Wald F
OVERALL MODEL	12.000	3474.892	0.000
MODEL MINUS			
INTERCEPT	11.000	53.633	0.000
INTERCEPT	.	.	.
PYR	1.000	250.133	0.000
INTWAGE	1.000	567.546	0.000
SEX	1.000	51.010	0.000
ED4CAT	3.000	1.864	0.151
RACECAT	3.000	11.963	0.000
MAR3CAT	2.000	34.351	0.000

Date: 05-18-2017
 Time: 14:27:37

SUDAAN

Page: 3
 Table: 1

Variance Estimation Method: Taylor Series (WR)
 SE Method: Robust (Binder, 1983)
 Working Correlations: Independent
 Link Function: Logit
 Response variable MDETV: MDETV
 For Subpopulation: PYR <= AGEONSETMDE
 by: Independent Variables and Effects.

Independent Variables and Effects	Odds Ratio	Lower 95% Limit OR	Upper 95% Limit OR
Intercept	0.032	0.023	0.045
PYR	1.033	1.029	1.038
INTWAGE	0.943	0.939	0.948
Sex 1=Male 2=Female			
1	0.641	0.565	0.727
2	1.000	1.000	1.000
Education 1=0-11 2=12 3=13-15 4=16+ Yrs			
1	1.000	1.000	1.000
2	0.980	0.858	1.120
3	1.097	0.977	1.232
4	0.981	0.863	1.114
Race 1=Other/Asian 2=Hisp/Mexican 3=Black 4=White			
1	1.000	1.000	1.000
2	0.780	0.594	1.024
3	0.633	0.468	0.857
4	1.077	0.848	1.367
Marital Status 1=Married 2=Previously Married 3=Never Married			
1	1.000	1.000	1.000

2	1.639	1.449	1.854
3	0.965	0.808	1.153
