

SAS Analysis Examples Replication C10

* SAS Analysis Examples Replication for ASDA 2nd Edition
* Berglund April 2017
* Chapter 10 ;

```
libname ncsr "P:\ASDA 2\Data sets\ncsr\" ;
```

```
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 ;
```

```
ods rtf style=normalprinter bodytitle ;  
title "Distribution of Age of Onset of MDE or Censor" ;
```

```
proc freq ;  
  tables ageonsetmde ;  
run ;  
title "Section 10.3.3 KM Example: Not Available in SAS SURVEY procedures but PROC LIFETEST will produce  
correct estimates but incorrect standard errors" ;
```

```
ods graphics on ;  
proc lifetest data=c10_ncsr outsurv=survival noprint ;  
title "Kaplan-Meier Survival Estimates by Race " ;  
time ageonsetmde*mde(0) ;  
strata racecat ;  
freq ncsrwtsh100 ;  
run ;  
ods graphics off ;  
proc format ;  
value rf 1='Other/Asian' 2='Hispanic/Mexican' 3='Black' 4='White' ;  
run ;
```

* Note, this graph is also produced by PROC LIFETEST but included here if a custom graphic is desired ;

```
legend1 label=('Race Group') frame ;  
proc sgplot data=survival ;  
series y=survival x=ageonsetmde / group=stratum markers ;  
yaxis min=0 max=1.00 ; xaxis min=0 max=100 ; *legend=legend1 vref=(.25 .5 .75) ;  
format stratum rf. ;  
title "Kaplan-Meier Survival Estimates" ;  
run ;
```

```
title " 10.4.5 Example: Fitting a Cox Proportional Hazards Model to Complex Sample Survey Data" ;  
proc surveyphreg data=c10_ncsr ;  
strata sestrat ; cluster seclustr ; weight ncsrwtsh ;  
class mar3cat (ref=first) sex (ref=last) ed4cat (ref=first) racecat (ref=first) / param=ref ;  
model ageonsetmde*mde(0) = intwage sex mar3cat ed4cat racecat ;  
run ;
```

* Note: PROC LIFETEST is used for parallel lines assumption test ;

```
ods graphics on ;  
proc lifetest data=c10_ncsr plots=(loglogs) noprint ;  
time Ageonsetmde*mde(0) ;  
strata racecat ;  
freq ncsrwtsh100 ;  
format racecat rf. ;  
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 ;
title "Print out of CASEID=1" ;
proc print ;
  where caseid=1 ;
  var caseid intwage ncsrwtsh sestrat seclustr pyr mdetv ageonsetmde ;
run ;
title "Discrete Time Logistic Regression: Table 10.5" ;
proc surveylogistic data=c10_expanded1 ;
  strata sestrat; cluster seclustr; weight ncsrwtsh;
  class racecat (ref=first) sex (ref=last) ed4cat (ref=first) mar3cat (ref=first) / param=ref;
  where pyr <= ageonsetmde ;
  model mdetv (event='1') =pyr intwage sex ed4cat racecat mar3cat / link=logit;
run;
title "Discrete Time Logistic Regression with Link=CLOGLOG: Table 10.6" ;
* Note that OR's are not produced automatically with LINK=cloglog, betas could be exponentiated
manually if desired ;
proc surveylogistic data=c10_expanded1 ;
  strata sestrat; cluster seclustr; weight ncsrwtsh;
  class racecat (ref=first) sex (ref=last) ed4cat (ref=first) mar3cat (ref=first) / param=ref;
  where pyr <= ageonsetmde ;
  model mdetv (event='1') =pyr intwage sex ed4cat racecat mar3cat / link=cloglog ;
  ods output parameterestimates=outest ;
run;
* obtain OR and CL manually ;
data outest_or ;
  set outest ;
  or=exp(estimate) ;
  upcl=exp(estimate + 1.96*stderr) ;
  lowcl=exp(estimate - 1.96*stderr) ;
run ;
title "Odds Ratios and CL for Table 10.6" ;
proc print data=outest_or ;
run ;
ods rtf close ;

```

Distribution of Age of Onset of MDE or Censor

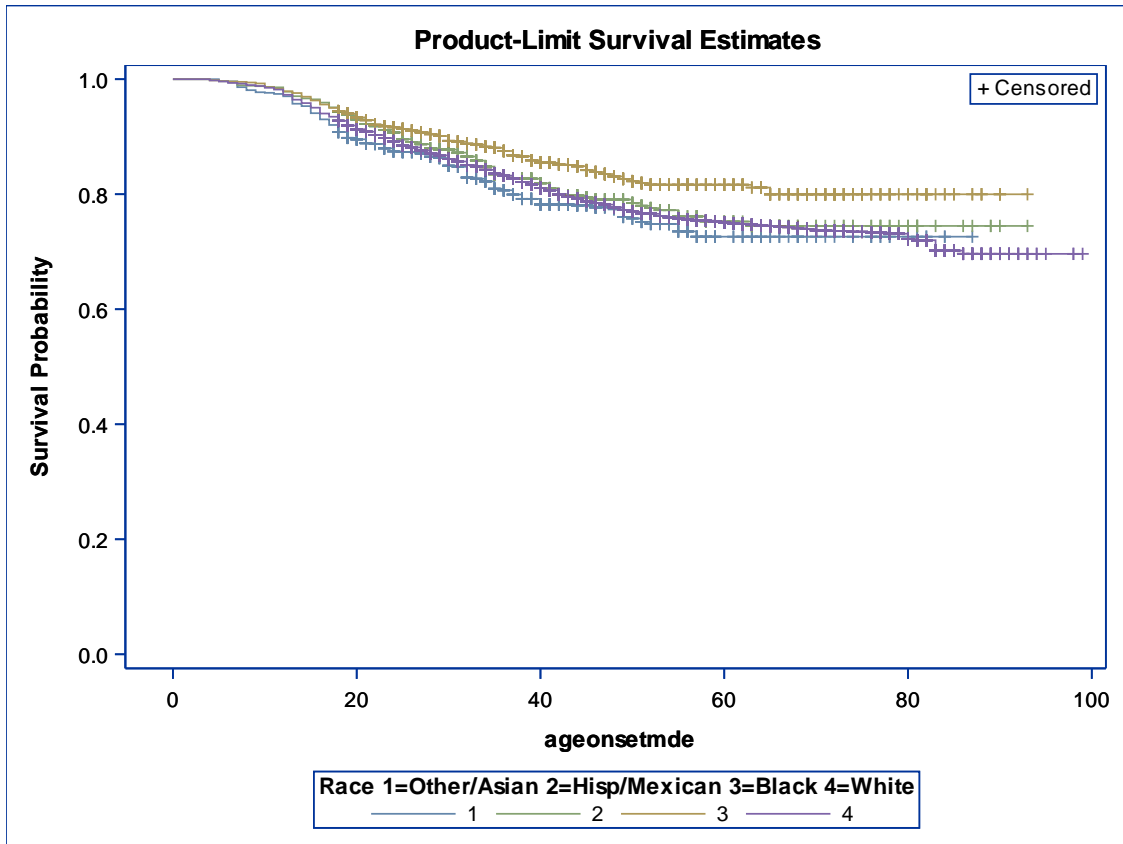
The FREQ Procedure

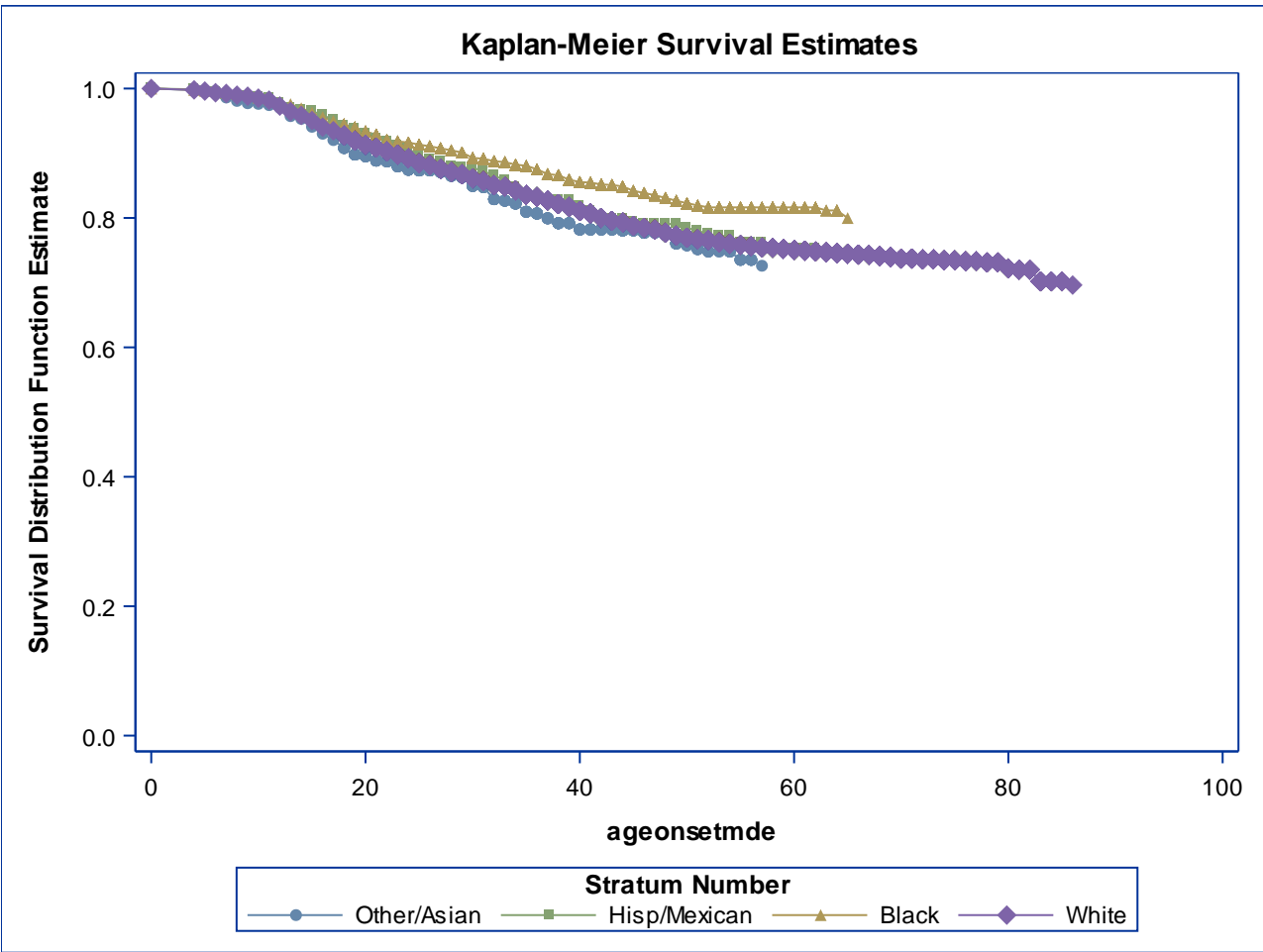
ageonsetmde	Frequency	Percent	Cumulative Frequency	Cumulative Percent
4	20	0.22	20	0.22
5	18	0.19	38	0.41
6	19	0.20	57	0.61
7	19	0.20	76	0.82
8	23	0.25	99	1.07
9	16	0.17	115	1.24
10	34	0.37	149	1.61
11	28	0.30	177	1.91
12	76	0.82	253	2.73
13	70	0.75	323	3.48
14	54	0.58	377	4.06
15	66	0.71	443	4.77
16	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

ageonsetmde	Frequency	Percent	Cumulative Frequency	Cumulative Percent
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
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

Kaplan-Meier Survival Estimates by Race

The LIFETEST Procedure





10.4.5 Example: Fitting a Cox Proportional Hazards Model to Complex Sample Survey Data

The SURVEYPHREG Procedure

Model Information		
Data Set	WORK.C10_NCSR	
Dependent Variable	ageonsetmde	
Censoring Variable	mde	Major Depressive Episode 1=Yes 0=No
Censoring Value(s)	0	
Weight Variable	NCSRWTSH	NCSR sample part 1 weight
Stratum Variable	SESTRAT	SAMPLING ERROR STRATUM
Cluster Variable	SECLUSTER	SAMPLING ERROR CLUSTER
Ties Handling	BRESLOW	

Number of Observations Read	9282
Number of Observations Used	9282
Sum of Weights Read	9282
Sum of Weights Used	9282

Design Summary	
Number of Strata	42
Number of Clusters	84

Class Level Information		
Class	Levels	Values
MAR3CAT	3	1 2 3
SEX	2	1 2
ED4CAT	4	1 2 3 4
racecat	4	1 2 3 4

Summary of the Number of Event and Censored Values			
Total	Event	Censored	Percent Censored
9282	1829	7453	80.30

Summary of the Weighted Number of Event and Censored Values			
Total	Event	Censored	Percent Censored
9282	1779.464	7502.536	80.83

Variance Estimation	
Method	Taylor Series

Model Fit Statistics		
Criterion	Without Covariates	With Covariates
-2 LOG L	31226.798	30366.511

Model Fit Statistics		
Criterion	Without Covariates	With Covariates
AIC	31226.798	30386.511

Testing Global Null Hypothesis: BETA=0				
Test	Test Statistic	Num DF	Den DF	p-Value
Likelihood Ratio	860.2864	10	Infy	<.0001
Wald	53.0216	10	33	<.0001

Analysis of Maximum Likelihood Estimates						
Parameter	DF	Estimate	Standard Error	t Value	Pr > t	Hazard Ratio
intwage	42	-0.049505	0.002380	-20.80	<.0001	0.952
SEX 1	42	-0.453581	0.062308	-7.28	<.0001	0.635
MAR3CAT 2	42	0.502499	0.060030	8.37	<.0001	1.653
MAR3CAT 3	42	0.080631	0.088835	0.91	0.3692	1.084
ED4CAT 2	42	-0.056883	0.067049	-0.85	0.4010	0.945
ED4CAT 3	42	0.045664	0.058017	0.79	0.4356	1.047
ED4CAT 4	42	-0.090199	0.063674	-1.42	0.1640	0.914
racecat 2	42	-0.249951	0.134372	-1.86	0.0699	0.779
racecat 3	42	-0.479031	0.148947	-3.22	0.0025	0.619
racecat 4	42	0.077782	0.117454	0.66	0.5114	1.081

Print out of CASEID=1

Obs	CASEID	intwage	NCSRWTSH	SESTRAT	SECLUSTR	pyr	mdetv	ageonsetmde
1	1	41	2.02426	1	2	1	0	34
2	1	41	2.02426	1	2	2	0	34
3	1	41	2.02426	1	2	3	0	34
4	1	41	2.02426	1	2	4	0	34
5	1	41	2.02426	1	2	5	0	34
6	1	41	2.02426	1	2	6	0	34
7	1	41	2.02426	1	2	7	0	34
8	1	41	2.02426	1	2	8	0	34
9	1	41	2.02426	1	2	9	0	34
10	1	41	2.02426	1	2	10	0	34
11	1	41	2.02426	1	2	11	0	34
12	1	41	2.02426	1	2	12	0	34
13	1	41	2.02426	1	2	13	0	34
14	1	41	2.02426	1	2	14	0	34
15	1	41	2.02426	1	2	15	0	34
16	1	41	2.02426	1	2	16	0	34
17	1	41	2.02426	1	2	17	0	34
18	1	41	2.02426	1	2	18	0	34
19	1	41	2.02426	1	2	19	0	34
20	1	41	2.02426	1	2	20	0	34
21	1	41	2.02426	1	2	21	0	34
22	1	41	2.02426	1	2	22	0	34
23	1	41	2.02426	1	2	23	0	34
24	1	41	2.02426	1	2	24	0	34
25	1	41	2.02426	1	2	25	0	34
26	1	41	2.02426	1	2	26	0	34
27	1	41	2.02426	1	2	27	0	34
28	1	41	2.02426	1	2	28	0	34
29	1	41	2.02426	1	2	29	0	34
30	1	41	2.02426	1	2	30	0	34
31	1	41	2.02426	1	2	31	0	34
32	1	41	2.02426	1	2	32	0	34
33	1	41	2.02426	1	2	33	0	34
34	1	41	2.02426	1	2	34	1	34
35	1	41	2.02426	1	2	35	0	34
36	1	41	2.02426	1	2	36	0	34
37	1	41	2.02426	1	2	37	0	34
38	1	41	2.02426	1	2	38	0	34
39	1	41	2.02426	1	2	39	0	34
40	1	41	2.02426	1	2	40	0	34
41	1	41	2.02426	1	2	41	0	34

Discrete Time Logistic Regression: Table 10.5

The SURVEYLOGISTIC Procedure

Model Information		
Data Set	WORK.C10_EXPANDED1	
Response Variable	mdetv	
Number of Response Levels	2	
Stratum Variable	SESTRAT	SAMPLING ERROR STRATUM
Number of Strata	42	
Cluster Variable	SECLUSTER	SAMPLING ERROR CLUSTER
Number of Clusters	84	
Weight Variable	NCSRWTSH	NCSR sample part 1 weight
Model	Binary Logit	
Optimization Technique	Fisher's Scoring	
Variance Adjustment	Degrees of Freedom (DF)	

Variance Estimation	
Method	Taylor Series
Variance Adjustment	Degrees of Freedom (DF)

Number of Observations Read	385696
Number of Observations Used	385696
Sum of Weights Read	386866
Sum of Weights Used	386866

Response Profile			
Ordered Value	mdetv	Total Frequency	Total Weight
1	0	383867	385086.58
2	1	1829	1779.46

Probability modeled is mdetv=1.

Class Level Information				
Class	Value	Design Variables		
racecat	1	0	0	0
	2	1	0	0
	3	0	1	0
	4	0	0	1
SEX	1	1		
	2	0		
ED4CAT	1	0	0	0
	2	1	0	0
	3	0	1	0
	4	0	0	1
MAR3CAT	1	0	0	
	2	1	0	

Class Level Information			
Class	Value	Design Variables	
	3	0	1

Model Convergence Status
Convergence criterion (GCONV=1E-8) satisfied.

Model Fit Statistics		
Criterion	Intercept Only	Intercept and Covariates
AIC	22706.046	21681.912
SC	22716.912	21812.302
-2 Log L	22704.046	21657.912

Testing Global Null Hypothesis: BETA=0				
Test	F Value	Num DF	Den DF	Pr > F
Likelihood Ratio	98.19	7.5762	318.20	<.0001
Score	49.47	11	32	<.0001
Wald	53.63	11	32	<.0001

NOTE: Second-order Rao-Scott design correction 0.4519 applied to the Likelihood Ratio test.

Type 3 Analysis of Effects				
Effect	F Value	Num DF	Den DF	Pr > F
pyr	250.13	1	42	<.0001
intwage	567.53	1	42	<.0001
SEX	51.01	1	42	<.0001
ED4CAT	1.86	3	40	0.1512
racecat	11.96	3	40	<.0001
MAR3CAT	34.35	2	41	<.0001

Analysis of Maximum Likelihood Estimates				
Parameter	Estimate	Standard Error	t Value	Pr > t
Intercept	-3.4355	0.1620	-21.21	<.0001
pyr	0.0328	0.00207	15.82	<.0001
intwage	-0.0583	0.00245	-23.82	<.0001
SEX 1	-0.4449	0.0623	-7.14	<.0001
ED4CAT 2	-0.0201	0.0661	-0.30	0.7622
ED4CAT 3	0.0929	0.0574	1.62	0.1133
ED4CAT 4	-0.0195	0.0633	-0.31	0.7603
racecat 2	-0.2484	0.1348	-1.84	0.0724
racecat 3	-0.4570	0.1499	-3.05	0.0040
racecat 4	0.0740	0.1182	0.63	0.5348
MAR3CAT 2	0.4942	0.0610	8.10	<.0001
MAR3CAT 3	-0.0353	0.0880	-0.40	0.6899

NOTE: The degrees of freedom for the t tests is 42.

Odds Ratio Estimates			
Effect	Point Estimate	95% Confidence Limits	
pyr	1.033	1.029	1.038
intwage	0.943	0.939	0.948
SEX 1 vs 2	0.641	0.565	0.727
ED4CAT 2 vs 1	0.980	0.858	1.120
ED4CAT 3 vs 1	1.097	0.977	1.232
ED4CAT 4 vs 1	0.981	0.863	1.114
racecat 2 vs 1	0.780	0.594	1.024
racecat 3 vs 1	0.633	0.468	0.857
racecat 4 vs 1	1.077	0.848	1.367
MAR3CAT 2 vs 1	1.639	1.449	1.854
MAR3CAT 3 vs 1	0.965	0.808	1.153
NOTE: The degrees of freedom in computing the confidence limits is 42.			

Association of Predicted Probabilities and Observed Responses			
Percent Concordant	63.6	Somers' D	0.406
Percent Discordant	23.1	Gamma	0.468
Percent Tied	13.3	Tau-a	0.004
Pairs	702092743	c	0.703

Discrete Time Logistic Regression with Link=CLOGLOG: Table 10.6

The SURVEYLOGISTIC Procedure

Model Information		
Data Set	WORK.C10_EXPANDED1	
Response Variable	mdetv	
Number of Response Levels	2	
Stratum Variable	SESTRAT	SAMPLING ERROR STRATUM
Number of Strata	42	
Cluster Variable	SECLUSTER	SAMPLING ERROR CLUSTER
Number of Clusters	84	
Weight Variable	NCSRWTSH	NCSR sample part 1 weight
Model	Binary Cloglog	
Optimization Technique	Fisher's Scoring	
Variance Adjustment	Degrees of Freedom (DF)	

Variance Estimation	
Method	Taylor Series
Variance Adjustment	Degrees of Freedom (DF)

Number of Observations Read	385696
Number of Observations Used	385696
Sum of Weights Read	386866
Sum of Weights Used	386866

Response Profile			
Ordered Value	mdetv	Total Frequency	Total Weight
1	0	383867	385086.58
2	1	1829	1779.46

Probability modeled is mdetv=1.

Class Level Information				
Class	Value	Design Variables		
racecat	1	0	0	0
	2	1	0	0
	3	0	1	0
	4	0	0	1
SEX	1	1		
	2	0		
ED4CAT	1	0	0	0
	2	1	0	0
	3	0	1	0
	4	0	0	1
MAR3CAT	1	0	0	
	2	1	0	

Class Level Information			
Class	Value	Design Variables	
	3	0	1

Model Convergence Status
Convergence criterion (GCONV=1E-8) satisfied.

Model Fit Statistics		
Criterion	Intercept Only	Intercept and Covariates
AIC	22706.046	21681.795
SC	22716.912	21812.185
-2 Log L	22704.046	21657.795

Testing Global Null Hypothesis: BETA=0				
Test	F Value	Num DF	Den DF	Pr > F
Likelihood Ratio	98.18	7.5810	318.40	<.0001
Score	49.47	11	32	<.0001
Wald	53.68	11	32	<.0001

NOTE: Second-order Rao-Scott design correction 0.4510 applied to the Likelihood Ratio test.

Type 3 Analysis of Effects				
Effect	F Value	Num DF	Den DF	Pr > F
pyr	250.29	1	42	<.0001
intwage	568.34	1	42	<.0001
SEX	50.97	1	42	<.0001
ED4CAT	1.85	3	40	0.1536
racecat	11.94	3	40	<.0001
MAR3CAT	34.45	2	41	<.0001

Analysis of Maximum Likelihood Estimates				
Parameter	Estimate	Standard Error	t Value	Pr > t
Intercept	-3.4444	0.1614	-21.34	<.0001
pyr	0.0327	0.00207	15.82	<.0001
intwage	-0.0582	0.00244	-23.84	<.0001
SEX 1	-0.4432	0.0621	-7.14	<.0001
ED4CAT 2	-0.0197	0.0659	-0.30	0.7658
ED4CAT 3	0.0924	0.0572	1.61	0.1139
ED4CAT 4	-0.0192	0.0631	-0.30	0.7623
racecat 2	-0.2474	0.1344	-1.84	0.0726
racecat 3	-0.4551	0.1494	-3.05	0.0040
racecat 4	0.0737	0.1179	0.63	0.5350
MAR3CAT 2	0.4928	0.0608	8.11	<.0001
MAR3CAT 3	-0.0355	0.0875	-0.41	0.6874

NOTE: The degrees of freedom for the t tests is 42.

Association of Predicted Probabilities and Observed Responses			
Percent Concordant	63.6	Somers' D	0.405
Percent Discordant	23.1	Gamma	0.467
Percent Tied	13.3	Tau-a	0.004
Pairs	702092743	c	0.703

Odds Ratios and CL for Table 10.6

Obs	Variable	ClassVal0	DF	Estimate	StdErr	WaldChiSq	ProbChiSq	tValue	ProbT	or	upcl	lowcl
1	Intercept		42	-3.4444	0.1614	455.565	0.00000	-21.34	<.0001	0.03192	0.04380	0.02327
2	pyr		42	0.0327	0.00207	250.285	0.00000	15.82	<.0001	1.03327	1.03747	1.02909
3	intwage		42	-0.0582	0.00244	568.335	0.00000	-23.84	<.0001	0.94348	0.94800	0.93898
4	SEX	1	42	-0.4432	0.0621	50.971	0.00000	-7.14	<.0001	0.64197	0.72503	0.56842
5	ED4CAT	2	42	-0.0197	0.0659	0.090	0.76434	-0.30	0.7658	0.98045	1.11553	0.86173
6	ED4CAT	3	42	0.0924	0.0572	2.607	0.10639	1.61	0.1139	1.09676	1.22687	0.98044
7	ED4CAT	4	42	-0.0192	0.0631	0.093	0.76084	-0.30	0.7623	0.98098	1.11012	0.86686
8	racecat	2	42	-0.2474	0.1344	3.391	0.06557	-1.84	0.0726	0.78081	1.01607	0.60002
9	racecat	3	42	-0.4551	0.1494	9.273	0.00233	-3.05	0.0040	0.63440	0.85030	0.47332
10	racecat	4	42	0.0737	0.1179	0.391	0.53164	0.63	0.5350	1.07652	1.35632	0.85444
11	MAR3CAT	2	42	0.4928	0.0608	65.763	0.00000	8.11	<.0001	1.63691	1.84397	1.45311
12	MAR3CAT	3	42	-0.0355	0.0875	0.164	0.68531	-0.41	0.6874	0.96515	1.14580	0.81298